

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 08:14:43 ; Search time 5416.12 seconds
(without alignments)
8658.665 Million cell updates/sec

Title: US-09-743-237-1

Perfect score: 2241

Sequence: 1 tctctgtgggttgcccg.....aaaaaaaaaaaaaaaaaaaaa 2241

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	ID	Description

1	2241	100.0	2241	10	MMU67176	U67176 Mus musculus
2	1747	78.0	1861	10	MMU77383	U77383 Mus musculus
3	1695.6	75.7	2215	10	AB057423	AB057423 Mus muscu
4	840.6	37.5	22856	10	AB057422	AB057422 Mus muscu
5	647.8	28.9	2134	9	HSU86074	UB6074 Homo sapien
6	169.8	7.6	179953	2	AC097959	AC097959 Rattus no
7	145.8	6.5	179953	2	AC097959	AC097959 Rattus no
8	141	6.3	157269	2	AP003096	AP003096 Homo sapi
9	123.6	5.5	157269	2	AP003096	AP003096 Homo sapi
10	107.4	4.8	188859	2	AP003732	AP003732 Homo sapi
11	85.4	3.8	214055	2	AC099385	AC099385 Rattus no
12	85.4	3.8	247253	2	AC103272	AC103272 Rattus no
13	85.4	3.8	297288	2	AC106533	AC106533 Rattus no
14	83.6	3.7	187360	9	AC021105	AC021105 Homo sapi
15	76.4	3.4	40893	3	CBRG45E19	AC084631 Caenorhab
16	74.8	3.3	100047	8	AC06081	AC006081 Arabidops
17	73.6	3.3	97714	8	ATF25024	AL078469 Arabidops
18	73.6	3.3	101679	8	ATF19B15	AL078470 Arabidops
19	73.6	3.3	198429	8	ATCHRIV70	AL161574 Arabidops
20	72.8	3.2	94362	8	AC005405	AC005405 BAC F18A1
21	72.8	3.2	170020	8	AP003256	AP003256 Oryza sat
22	72.8	3.2	180206	2	AP003274	AP003274 Oryza sat
23	71.2	3.2	39553	3	CEJCB	Z82274 Caenorhabdi
24	71.2	3.2	330724	2	CEY67H2	AL022475 Caenorhab
25	69	3.1	27555	2	AC012903	AC012903 Drosophil
26	69	3.1	146153	3	AC007452	AC007452 Drosophil
27	69	3.1	159065	3	AC01634	AC091634 Drosophil
28	69	3.1	264646	3	AE003818	AE003818 Drosophil
29	67.6	3.0	188859	2	AP003732	AP003732 Homo sapi
30	63.2	2.8	128463	2	AP003619	AP003619 Oryza sat
31	63.2	2.8	159947	2	AP003763	AP003763 Oryza sat
32	63	2.8	188844	2	AC099384	AC099384 Rattus no
33	61.4	2.7	64971	2	AC107744	AC107744 Mus muscu
34	53.6	2.4	2146	8	AF205142	AF205142 Arabidops
35	53.4	2.4	125020	9	AF429315	AF429315 Homo sapi
36	53.2	2.4	79355	8	AB012247	AB012247 Arabidops
37	53	2.4	1793	8	AY046019	AY046019 Arabidops
38	53	2.4	2298	8	AF204059	AF204059 Arabidops
39	53	2.4	2431	8	AF206324	AF206324 Arabidops
40	50.8	2.3	2110	5	AF078111	AF078111 Xenopus l
41	50.8	2.3	188844	2	AC099384	AC099384 Rattus no
42	49.6	2.2	158782	2	AC044874	AC044874 Homo sapi
43	49.2	2.2	1072	3	AB024943	AB024943 Tegula pf
44	47.2	2.1	154242	2	AC073373	AC073373 Mus muscu
45	47	2.1	163788	2	AC018996	AC018996 Homo sapi

ALIGNMENTS

RESULT 1

MMU67176

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

MMU67176 2241 bp mRNA linear ROD 13-APR-1999
Mus musculus tesmin-1 mRNA, complete cds.
U67176
U67176.1 GI:4581558

house mouse.
Mus musculus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 2241)

Sugihara,T., Wadhwa,R., Kaul,S.C. and Mitsui,Y.

A novel testis-specific metallothionein-like protein, tesmin, is an
early marker of male germ cell differentiation

Genomics 57 (1), 130-136 (1999)

99208669

2 (bases 1 to 2241)

Sugihara,T.

Direct Submission

Submitted (19-AUG-1996) NIBH, AIST, Higashi 1-1, Tsukuba, Ibaraki,

Japan, 305

Location/Qualifiers

```

source      1. .2241
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
651. .1538
/codon_start=1
/product="tesmin-1"
/protein_id="AAD24666.1"
/db_xref="GI:4581559"
/translacion="MVICQLKGGAOMLICDNCARELKAHLHLLPOYDQDSFFQSELP
KPMITLAVGRLLPVPKALNLTLOVDNGALPSAVNGAAPPSPALQGPKPITLSGYCDDCF
SSGDFCNSCSNNLRHLEERFKAIKACLDNPEAFQPMGKRLGRLAKLRLKSGCNCX
RSGDLKNCYEAIKIMCSICACIKADKNYESPERKMLMSTPHYMEPGDFESSHYLS
PAKFSPPKLRKNRQAFSCISWEVEATCACLLAQGEAEQEHCSPSLAQOMILEBFFG
RCLSQIHIEBFKSKGLKIE"

BASE COUNT      558 a 554 c 583 g 546 t
ORIGIN

Query Match      100.0%; Score 2241; DB 10; Length 2241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 tATCCTGtgggttggccccggcgagggctctcagcaggctcagccacaaggagata 60
Db 1 TATCCTGTGGTGTGGCCCCGGGACAGAGCTCTCAGCAGGCTCAGGACCAACAAGGATA 60

Qy 61 cAcagtggttctcctggcctgttgagcttgactccaccacactccgcccagcagg 120
Db 61 CACAGTGTGTGTCTTGCCCTGTGGACTTGTGACTCCACCACCTCGGCCCCAGCAGG 120

Qy 121 ctAgggatagaacccaggcgcttttgogtgtctgcagatagttctcagcctggtagtt 180
Db 121 CTAGGGATAGAACCCAGGCGCTTTTGGCTGTCTGCAGATAGTCTTCAGCCTGGTAGTT 180

Qy 181 ggggttgctgggagatttttttcttccacacaaagacttccattaggagatttt 240
Db 181 GGGGTGGCTGGGAGATTTTTTTTCTTCACACAAAGACTTCCATTATTGAGGATTTT 240

Qy 241 tcaGttaatgatctccccctctgaagataaggagacagttctttaacctatgagct 300
Db 241 TCAGTTGATGATCTCCCCCTCTGTAAGTAAGGAGCAGTGTCTTTAACTATGTAGAGT 300

Qy 301 ttGatgaattctgtctctcaaccatatgtctaagctatatagcaattccttgaattgc 360
Db 301 TTTGATGAATTCGTCTCTCAACCATATGCTAAGCTATATAGCAATTCCTTGAATTGC 360

Qy 361 tAtaacttaggagaacctgattctcctgcctctacatcctgtagtgcgtagtgata 420
Db 361 TATATACTTAGGAGAACCCTGATCTCTCTGCCTCTACATCCTGAGTGTAGTGATACA 420

Qy 421 gggggaatacatttggtagactccgatgaactgccaactgccaaggttccagaggcaga 480
Db 421 GGGGGAATCAATTTTGGTAGACTCCGATGAACCTACTGCGAGGTTCCCAAGGACGCAAG 480

Qy 481 aagcaagaaaagtgttgaaatcaagaaagcagtggttagtgccagggcgagccctg 540
Db 481 AAGCAAGAAAAGTGTGAATCAAAAGCAGGTGGTAGTGCCAGGGCGGAGCCCTG 540

Qy 541 aagcgcagctttccaggccctctggctcaggaatcctgttgcagattccccatcccc 600
Db 541 AAGCGCAGCTTTCCAGGCGCCTCTGGCTCAGGAATCCTGTGTGAAGTTCCCATCATCC 600

Qy 601 aggaggcagaggagcctccagctgcctcggaagaagactccagcccattggtgattt 660
Db 601 AGGAGGAGGAGGAGGCTCCAGCTGCGCCTCGGAAGAAGACTCCAGCCCCATGGTGATT 660

Qy 661 gtCagctgaagaggagcgccagatgctctgcatagacaaactgtggcgagaggagctca 720
Db 661 GTCAGCTGAAGGAGGCGCCAGATGCTCTGCATAGACAACTGTGGCGCGAGGAGCTCA 720

Qy 721 aagcgctccatctgtcttctcagtcagatgaccagagcagtttccctcagtcagagctcc 780

```



```

QY 1861 ttattatgagccagctcaagtttagagagctaggctcttcttcaggtggagctctgcccaa 1920
Db 1861 TTATTATGAGCCAGCCTCAAGTTAGAGAGCTAGGCTCTTCTCAGGTGGACTCTGCCCAA 1920
QY 1921 atcacatacaagtcaggtggccatcagggggtttttccagggccagggctctgacagagat 1980
Db 1921 ATCACATACAAAGTCAGGTGGCCATCAGGGGTTTTTCCAGGCCAGGCCCTGTGACAGGAGAT 1980
QY 1981 atgggagggggtcgaggttagaagctgggtttgtttggatttttttggatttttttttccct 2040
Db 1981 ATGGGAGGGGGGTCGGGTTAGAGCTGGGTGTGTTGGATTTTTTGGGTTTTTTCCTCTCCT 2040
QY 2041 gtattctgtctgaagtgaagaaactgtctctctgtctccaaacttttctccataattactg 2100
Db 2041 GTATTCTGCTTGAAGTGAGAAACTTGCTCTCTGCTGCCAACCCTTTCTCCATAATTAAGT 2100
QY 2101 ctgacaggtcgctgtgacctgacagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc 2160
Db 2101 CTGACAGGTGCGCTGTGACAGTCACAGTGACCTCAGACACAGAGAGTGAGGTGGCTT 2160
QY 2161 attatgcccaacttgggtttgtttgtgtgagataaacctttccagactcccaaaaaaaa 2220
Db 2161 ATTATGCCACACTTGTGTTTGTGTGAGATAAAACCTTTCCAGACTCCCAAAAAAAA 2220
QY 2221 aaaaaaaaaaaaaaaaaa 2241
Db 2221 AAAAAAAAAAAAAAAAAA 2241

RESULT 2
LOCUS MMU77383 1861 bp mRNA linear ROD 13-APR-1999
DEFINITION Mus musculus tesmin-2 mRNA, complete cds.
ACCESSION U77383
VERSION 077383.1 GI:4581560
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 1861)
AUTHORS Sugihara,T., Wadhwani,R., Kaul,S.C. and Mitsui,Y.
TITLE A novel testis-specific metallothionein-like protein, tesmin, is an
early marker of male germ cell differentiation
JOURNAL Genomics 57 (1), 130-136 (1999)
MEDLINE 99208669
REFERENCE 2 (bases 1 to 1861)
AUTHORS Sugihara,T.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-1996) NIBH, AIST, Higashil-1, Tsukuba, Ibaraki
305, Japan

FEATURES
source Location/Qualifiers
1..1861
/organism="Mus musculus"
/strain="ICR"
/db_xref="taxon:10090"
271..1158
/codon_start=1
/product="tesmin-2"
/protein_id="AAD24667.1"
/db_xref="GI:4581561"
/translation="MVICOLKGAOMLIDNCGARELKALHLLLPQYDDOSSPQSEL
KPTVLVGLLPPAKLNLITQVDNGLPSAVNGAAPPSPALQGPPLKILSGYCD
SSGDFNSCSNNLRELEFRKAIKACLDNRNPEAFQPKMGKRLGAKLRHSGCNCK
RSGCLNKYCEYAKIMCSICKTACKNYEESPKMLMSTPHYMEPDGFSSHYLS
PAKFGPPLKRNQAFKISWVEVETACALLAQGEAEQEHCSPLAEQMLIEBFG
RCLSQLHTEFKSGKLE"
BASE COUNT 456 a 491 c 498 g 426 t
ORIGIN

Query Match 78.0%; Score 1747; DB 10; Length 1861;

```

```

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 gttgaatacaaaagaagcaggtgtgttagtgccagggcgagccctgaagcagcttc 554
Db 115 GTTGAATACAAAAGAAGCAGGTGTGTGTGTCGCCAGGGCGAGCCCTGAAGACGCGAGCTTC 174
QY 555 cagggccctctgctcaggaatcctgttcaagttcccatccatccagagagagagagag 614
Db 175 CAGGCCCTCTGCTCAGGAATCTGTGTGAAGTTCCCATCATCCAGAGGCGAGAGAG 234
QY 615 gctccagctgcccctcggaagaagactccagccccatggtgatttttcagctgaaagga 674
Db 235 GCCTCCAGCTGCCCTCGGAAGAAGACTCCAGCCCCATGTTGATTTCTCAGCTCAAAAGGA 294
QY 675 ggcgcccagatgctctgcatagacaactgtggcgaggaggagctcaaaagcgtccactg 734
Db 295 GGCGCCAGATGCTCTGCATAGACAACCTGTGGCGGAGGAGCTCAAAAGCGCTCCATCTG 354
QY 735 ctctcagtaacgatgaccagagcagtttccctcagtcagagctccctaaagccaatgaca 794
Db 355 CTCTCTCAGTACGATGACAGAGAGCGATTTCCCTCAGTCAGAGCTCCCTAAGCCCAATGACA 414
QY 795 actttagtgggaagactctctgcccagtagaccagcgaagttaaatctcatcacacagttgat 854
Db 415 ACTTTAGTGGGAAGACTTCTGCCAGTACCAGGAAGTTAAATCTCATCACACAGGTTGAT 474
QY 855 aatggagctctcccatcagctgtcaatggggcgctcttccctctgagacctgctctgcaa 914
Db 475 AATGGAGCTCTCCCATCAGCTGCAATGGGGCTGCCCTTCCCTCTGGACCTGCTCTGCAA 534
QY 915 gggccaccccaataactctgtggtactgtgactgtacttccagcgggagctctgc 974
Db 535 GGGCCACCCCAATAACTCTGTCTGGTACTGTGACTGTCTCCAGCGGGAGCTTCTGCG 594
QY 975 aacagctgcagctgcacaacacctgcgcatgagctcagcgtctcaaaagccataaaagcg 1034
Db 595 AACAGCTGCAGCTGCACAACCTGCGCCATGAGCTCAGCGCTTCAAAAGCCATAAAGCG 654
QY 1035 tgcctgataagaatcctcgaagctttccaccccaaaaaaaggggaaagccgtctggagct 1094
Db 655 TGCTTGTATGAATACTCTGAAGCTTTTCAACCAAAAAATGGGAAAGCCGCTCTGGGAGCT 714
QY 1095 gctaaacttcgacacagcagaaggtgcaactgaagcgtcagcgtccctgaaagaaactac 1154
Db 715 GCTAAACTTCGACACACAAAGGGTGCAACTGTGAAGCGCTCAGGCTGCCCTGAAAGAACTAC 774
QY 1155 tgtgagtgtatgaggccaaaatcattgttcttccatttgcaaatgcatgtgtgcaaa 1214
Db 775 TGTGAGTGTATGAGGCCAAAATCATCTGTTCTTCCATTTTGCAAAATGCATTGCTTGCAA 834
QY 1215 aactagaagaagtcagacacgaataatgctgatggagcacaccccccactacatggagcct 1274
Db 835 AACTATGAAGAAAGTCCAGAACCGAAAAATGCTGATGAGCACACCCCACTACATGAGACCT 894
QY 1275 ggggacttcagagcagccattatttcccccagccaaagttctcagagctcccaaaactg 1334
Db 895 GGGGACTTTCAGAGCAGCCCATTTTGTCCCGCCCAAGTTCTCAGGACCTTCCAAAACCTG 954
QY 1335 agaaaaaataggcaggccttctctctgtatctctctgggaagtagtggaggccacatgtgcc 1394
Db 955 AGAAAAAATAGGCAGGCTTCTCTCTGTATCTCTGGGAAGTAGTGGAGGCCACATGTGCC 1014
QY 1395 tgcctgtgcccaggtgaggaaagcagagagcagtcagtcagtcagtcagtcagtcagtcag 1454
Db 1015 TGCTGTGTCGCCAGGTGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1074
QY 1455 caatgactctcagagaggtttgaaagtgctgcagattctccacatccagcttcaag 1514
Db 1075 CAGATGATCTCTGGAGGAGTTTGGAGGTGCTGTCCAGAGATTCTCCACATCGAGTTCAAG 1134
QY 1515 tcaaggggctgaaaattgagtagcgtgcaagctggtaaaaggggaatgcctgtgagcagc 1574

```

Db	1135	TCCAAGGGCTGAAAATTGAGTACGCTGCACAGCTGGTAAAGGGGAATCCCTGTGGC AAGC	1194
Qy	1575	ctcagccctggaaatctgcacccgaggaaagctggcgccagggagagcagagggccgcgca	1634
Db	1195	CTCAGCCCTGGGAATCTGCACCGAGGAAGTGGTGCCAGGGAGGAGCAGAGGGCCGGCA	1254
Qy	1635	tcatggccaggtcagctgtaggtctgagtgatctgcatggttactggccagcctactcaa	1694
Db	1255	TCATGGCCAGGTCAGCTGTGAGGTCTGAGTGTGATCTGATGTGCTACTGGCCAGCCTACTCAA	1314
Qy	1695	ggatctctaaagtgcacaagcaggcagagccaccctggggatgggaatggacaactggccctcctgtcc	1754
Db	1315	GGTATCTAAAGTCAACAGCAGCAGCCACCCCTGGGGATGGACACTGGCCCTCTGTGCC	1374
Qy	1755	ctggggagggccctctggggactccctccctgcataaaaaagagggtgaattttctacttgt	1814
Db	1375	CTGGGAGGCCCTCTGGGGACTCCCTGCCCTGCATAAAAGAGGGGTGATTTCCTACTTGT	1434
Qy	1815	tgttatgttttgctttccaaattgcttagtagtacctccatccaagtattattatgagccag	1874
Db	1435	TGTTATGTGTGCTTTCAAATGCTTAGTAGTACCTCCATTCAAGTTATTATGAGCCAG	1494
Qy	1875	cctcaagttagagcgtaggtctcttctcaggtggagctctgccaaatcacatacaagaagtc	1934
Db	1495	CCTCAAGTTAGAGAGCTAGGCTCTTCTTCAGTGGGAGCTGTGCCCAAAATCACATACAAGTC	1554
Qy	1935	agtgggccatcagggggttttccagggccagcctctgtacaggagatatggagggggggtc	1994
Db	1555	AGTGGCCATCAGGGGTTTTTCAGGGCCAGCCCTGTGACAGGAGATATGGGAGGGGGGTC	1614
Qy	1995	gggttagagctgggtttgtttggatttttgcgttttttcttctctgtatttctgtctga	2054
Db	1615	GGTTTAGAGCTGGGTGTTGTAATTTTTCGCTTTTTCCTCCTGCTATTTCCTGCTTGA	1674
Qy	2055	agtgagaaaactgtctcctgtccaaactttctccataattactgtcagcgctgcct	2114
Db	1675	AGTGAGAAAACTTCTCTCTGTCCAACTTTCTCCATAAATTACTGTGCAAGTCTGCCT	1734
Qy	2115	gctgaccagtcacagtgacctcagacccagaaggttaggtggtcttatatgccacact	2174
Db	1735	GCTGACCAGTACAGTGACCTTGACACACCAAGAGGTTGAGGTGGGTATTATATGCCACACT	1794
Qy	2175	ttgtgtttgtgtgagaaataaacctttccagactcccaaaaaaataaaaaaataaaaa	2234
Db	1795	TTGTGTTTTGTTGTGAGAAATAACCTTTCCAGACTCCCAAAAAAAAAAAAAAAAAA	1854
Qy	2235	aaaaaa 2241	
Db	1855	AAAAAA 1861	

RESULT	3	
LOCUS	AB057423	
DEFINITION	AB057423	2215 bp mRNA linear ROD 26-DEC-2001
ACCESSION	AB057423	Mus musculus mRNA for tesmin, complete cds, strain: BALB/c.
VERSION	AB057423.1	GI:15617421
KEYWORDS		
SOURCE		Mus musculus (strain: BALB/c) tissue_lib:testis cDNA to mRNA.
ORGANISM		Mus musculus
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (sites)	
AUTHORS	Sutou, S., Miwa, K., Matsuura, T., Kawasaki, Y., Ohinata, Y. and Mitsui, Y.	
TITLE	Structure of the tesmin gene encoding a testis-specific persistent protein; a possible multifunctional protein with dynamic changes of localization throughout spermatogenesis	
JOURNAL	unpublished	
REFERENCE	2 (bases 1 to 2215)	
AUTHORS	Sutou, S. and Mitsui, Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-MAR-2001) Shizuyo Sutou, National Institute of	

Advanced Industrial Science and Technology (AIST), Institute of Molecular & Cell Biology; Tsukuba Central 6, Higashi 1-1-1, Tsukuba, Ibaraki 305-8566, Japan (E-mail: s-suto@aist.go.jp, Tel: 81-298-61-6052, Fax: 81-298-61-9498)

Location/Qualifiers

1. .2215

/organism="Mus musculus"

/strain="BALB/c"

/db_xref="taxon:10090"

. /tissue_lib="testis"

82. .1509

/gene="tesm"

82. .1509

/gene="tesm"

/codon_start=1

/product="tesmin"

/protein_id="BAB64935.1"

/db_xref="GI:115617422"

/translation="MEDALLIGAMTPEDELGAELFSERVFAFDLALSPAGGAADRDVLPVLADAYLCATGPEVPPALALGDFPGLPELRSPDDAAQAPAPPAVXSVHVLSSLPGARGPALLPLSGARVIVPEIKEAGSVPGGSPEDAATQAPLAQSCCKFVPSQAEAEASCCRBDSSDPWVLCOLKGQMLCIDNCGARLELKALHLLPQYDDQSFQPSQELPMTHLVRLPLVPKLVNLTVDNGALPSAVNGAAPPGLQPPKPTILISGYDCDFSGDGNCSNNLRHELERFKAIKADLNREFAQPMKMGKRGILAKLRHSKGCNKRSGCLKNYCSEYAEKIMCSSTICKIACKNIEESPERKMLSTPHYMPGDFESSHYLSPAPFSGPPKLRKNQAFSCISWEVVEATCACLQAEGEAEQHCSPSLAEOMILIEFRCGLSLIHLIEFKSKGLIE"

BASE COUNT 497 a 607 c 635 g 476 t

ORIGIN

Query Match 75.7%; Score 1695.6; DB 10; Length 2215;

Best Local Similarity 99.5%; Pred. No. 0;

Matches 1742; Conservative 0; Mismatches 4; Indels 4; Gaps 4;

Qy 495 gttgaaatcaagaagcaggtgtagtbtgccagggcgagccctgaagcgcagcttc 554

Db 466 GTTGAAATCAAGAAGCAGGTGTGTGTGTCGACGGCGCAGCCCTGAAGACGCAGCTTC 525

Qy 555 caggccctctggtcctgaagtcctgttgaagttccatccatccaggagcagagag 614

Db 526 CAGGCCCCCTTGGCTCAGGAATCTGTGTGAAGTTCCTCATCCCCAGGAGCGAGAGG 585

Qy 615 gcctccagctgccctcggaagaagactccagccccatggtattgtcagctgaagga 674

Db 586 GCCTCCAGCTGCCCTCGAGAGAAGACTCCAGCCCCATGGTATTGTTCAGCTCAAAGGA 645

Qy 675 ggogccagatgtctctgcatagacaactgttgogcgaggagctcaaagcgctccatctg 734

Db 646 GGCSCCCAGATGCTCTGCATAGACAACGTGTGGCGGAGGGAGCTCAAAGGCGCTCCATCTG 705

Qy 735 ctctcctcagtaacgatgaccagagcagtttccctcagtcagagctccctaagccaataca 794

Db 706 CTTCTCAGTACGATGACCAAGAGAGCTTTCCCTCAGTCAGAGCTCCCTTAAGCCAAATGACA 765

Qy 795 accttagtgggaagactcttcgcagtagcaccgcaagttaaattcctcatcacacaggttgat 854

Db 766 ACTTTAGTGGAGAGACTTCTGCCAGTACCAGCGAAGTTAAATCTCANTCACACAGTTGAT 825

Qy 855 aatggagctctcccatcagctgttcaatggggtgcctttccctctggacctgctctgcaa 914

Db 826 AATGGAGCTCTCCATCAGCTGTCAATGGGGCTGCCCTTCCCTCTGGACCTGCTCTGCAA 885

Qy 915 ggcccaaccataaactctgtctgggtactgtgaactgtcttccctccagcggggacctctgc 974

Db 886 GGCCACCCCAAAATAACTCTGTCTGGGTACTGTGACTGTCTCTCCACGGGGACTTCTGC 945

Qy 975 aacagctgcagctgcacaacacctgcgcctatgagctgcagcgcttcaagcaccataaagcg 1034

Db 946 AACAGCTGCAGCTGCACAACCTGGCGCCATGAGCTCGAGCGCTTCAAGGCCATAAAGCG 1005

Qy 1035 tgtcttgatgaataacctgaagcttttccaccacaaaaatggggaagggccgtctctgggagct 1094

```
Db 1006 TGTCTTGATAGAAATCTGAAGCTTTCCAAACCAAAAAATGGGAAAGGCCCTCTCGGAGCT 1065
Qy 1095 gctaaactcgacacagcaagagtgcaactgtaagcgctcagctgctccgtaagaactac 1154
Db 1066 GCTAAACTTCGACACACAAAGGTTGCACTGTAAAGCGCTCAGGCTGCCGTGAAGACTAC 1125
Qy 1155 tgtgagtgctatgaggccaaatcatgtgttcttccatttgcaaatgattgcttgcctgcaaa 1214
Db 1126 TGTGAGTGTATGAGGCCAAATCATGTGTTCTTCCATTTGCCAAATGCAATGCTTGCCTG 1185
Qy 1215 aactatgaagaagtcagaaacgaaaaatgctgatgagcacacccccactacatgagcct 1274
Db 1186 AACTATGAAGAAAGTCCAGAACGAAAAATGCTGATGAGCACACACCCCTACTACATGAGCCT 1245
Qy 1275 ggggacttgaagcagcattatttccccagcccaagttctcagacacctcccaaaactg 1334
Db 1246 GGGACATTTGAGACAGCCCATTTTGTCCCCAGCCCAAGTTCTCAGGACCTTCCAAAAGTG 1305
Qy 1335 agaaaaataggcgggcttctctgtatctctctggaagtagtgaggggccacatgtgcc 1394
Db 1306 AGAAAAATAGGCAGGCTTCTCTGTATCTCTGGGAAGTAGTGAGGCCACATGTGCC 1365
Qy 1395 tgcctgtgccccgggtgagggaagcagagcagagcactgttccccaaagcttgctgag 1454
Db 1366 TGCTGTGCCCCAGGTTGAGGAAGCAGAGCAGAGCAGTGTTCCTCCCAAGCTTGGCTGAG 1425
Qy 1455 cagatgactcggagagtttggaagtgctgtcgcagagatttccacatcagagattcaag 1514
Db 1426 CAGATGATCTCGAGGAGTTTGGGAAGTGCTGTGCGAGAGATTCTCCACATCGAGTTCAAG 1485
Qy 1515 tccaggggctgaaaaattgagtagctgcaagctggcaaggggaatgctgtggtgcaagc 1574
Db 1486 TCCAAGGGCTGAAAAATGAGTAGCTGCAAGCTGTGTAAAGGGGAATGCCGTGGCAAGC 1545
Qy 1575 ctccgacctgggaatctgcacccgggaagctggtgccccagggagagcagagccgcgca 1634
Db 1546 CTCAGCCCTGGGAATCTGCACCGAGGAAGCTGTGTGCCAGGAGGAGCAGAGGCGCGCA 1605
Qy 1635 tcatggcaggtcagctgtaggtgagtgatctgcatgtagtactgcccagctactcaa 1694
Db 1606 TCATGGCCAGGTGACGTGTGAGGTCTGAGTGATCTGATGTTGATGTCAGGACCTTACTCAA 1665
Qy 1695 ggatctcaagtgcagcagcagcagcagccctgggagtggaactggcactggccctcctgtcc 1754
Db 1666 GGTATCTAAAGTCAAGCAGCAGCAGCAGCCCTGGGGATGGACACTGGCCCTCTCTGCC 1725
Qy 1755 ctggggagggcctctggggactccctgcccgtgataaaagaggggtgatttctacttgt 1814
Db 1726 CTGGGAGGCGCTCTGGGGACTCCCTGCGCTGCATATAAAAGAGGGGTATTTTCTACTGT 1785
Qy 1815 tgttatgttcttcaaatgtcttagtagtacctccattcagttattatgagcag 1874
Db 1786 TGTTATGTGTTGTTTCAAAATGCTTAGTAGTACCTTCCATTCAGTTATTATGAGCCAG 1845
Qy 1875 cctcaagttagagcagctgctcttctcagttgagactctgcccacaaatccacatacaagtc 1934
Db 1846 CCTCAAGTTAGAGAGCTAGGCTCTTCTCAGGTGGACTCTGCCCAATTCATACATACAGTC 1905
Qy 1935 aggtggccatcaggggtttttccagggcagggcctctgacagggagatagggga-ggggggg 1993
Db 1906 AGGTGGCCATCAGGGG-TCTTCCAGGCCAGGCCCTGTGCACAGGGGATATGGAGGGGGGT 1964
Qy 1994 cgggttagagctgggttggattttttgg-ttttttctctctgtatttctctt 2052
Db 1965 CGGCTTAGCTGGGTGTTTGGATTTTGGCTTTTCTCTCTGTTATTTCTGCTT 2024
Qy 2053 gaagtgaagaaactgtctctgtccaa-ccttttctccataaataactgtgcacggctg 2111
Db 2025 GAAGTGAGAAAACTGTCTCTGTGCCAACCTCTGCTCCATAAATTAAGTGTGACGGTGC 2084
Qy 2112 cctgctgaccagtcacagtgacacctcagacaccagagaggtgaggttggtattatgcccac 2171
Db 2085 CCTGCTGACCAGTACAGTGAACCTTCAGACACACAGAAAGGTGAGGTGGGTATTATGCCCC 2144
```

```
Qy 2172 acttgtgtttgtgtgagaataaacctttccagactccccaaaaaataaaaaa 2231
Db 2145 ACTTTGTGTTTGTGTGAGANTAACTTTCCAGACTCCCAAAAAAAAAAAAAA 2204
Qy 2232 aaaaaa 2241
Db 2205 AAAAAA 2214

RESULT 4
AB057422 22856 bp DNA linear ROD 26-DEC-2001
LOCUS Mus musculus gene for tesmin, complete cds, strain:129.
DEFINITION AB057422
ACCESSION AB057422
VERSION AB057422.1 GI:15617419
KEYWORDS
SOURCE Mus musculus (strain:129) DNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (sites)
Sutou, S., Miwa, K., Matsuura, T., Kawasaki, Y., Ohinata, Y. and
Mitsui, Y.
REFERENCE
AUTHORS Structure of the tesmin gene encoding a testis-specific persistent
protein: a possible multifunctional protein with dynamic changes of
localization throughout spermatogenesis
Unpublished
JOURNAL 2 (bases 1 to 22856)
REFERENCE Sutou, S. and Mitsui, Y.
AUTHORS Direct Submission
TITLE Submitted (13-MAR-2001) Shizuho Sutou, National Institute of
JOURNAL Advanced Industrial Science and Technology (AIST), Institute of
Molecular & Cell Biology, Tsukuba Central 6, Higashi 1-1-1,
Tsukuba, Ibaraki 305-8566, Japan (E-mail: s-sutou@aist.go.jp,
Tel: 81-298-61-6052, Fax: 81-298-61-9498)
FEATURES
Location/Qualifiers
source 1..22856
/organism="Mus musculus"
/strain="129"
/db_xref="taxon:10090"
gene join(2569..2952,5717..5875,7539..7659,9534..9607,
10577..10668,16100..16190,17665..17802,18550..18725,
20538..20730)
/genes="tesm"
CDS join(2569..2952,5717..5875,7539..7659,9534..9607,
10577..10668,16100..16190,17665..17802,18550..18725,
20538..20730)
/genes="tesm"
/codon_start=1
/product="tesmin"
/protein_id="BAB64934.1"
/db_xref="GI:15617420"
/translation="MEDALLGAMTGPDELGAELFSGSERVAFADGLALSPAGGAARDRE
LPVLADYLGATEPPEPLRALSPVPAEKPAAALGDFPGLPELRSPDDAAPVAVSV
HVLSSLLPGAGPALLPLSAGVRVPIVEIKAGSVPGGSPEDAAFOAPLAQSCCKF
PSQGEAEASCCPRKKDSSPWICOLKGGQMLCINDCGARELKALHLLPOYDDQSF
PQSLDFKPLMTLLGRLLPVPKLNLTITQVNDGALPNSAVNGAAPPSPGALOGPKTILS
GYCDLFSGGDFCNSCCNLRHELEFKAIKACLDNRNPEAFQPMKMGRLGAKLRHS
KGCNKRSGCKNYCEYEAKIMCSICTACKNYEESPERKMLMSTPHYMEPDPFE
SHYLSPAKFSFGPKLNKQAFSCISWEVVEATCACLALHGBEAHEHCSPSLAEOM
ILBEFGRLSOLHIEFKSKGLKIE"
```

```
BASE COUNT 5724 a 5667 c 5552 g 5913 t
ORIGIN

Query Match 37.5%; Score 840.6; DB 10; Length 22856;
Best Local Similarity 99.3%; Pred. No. 3.4e-244;
Matches 865; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
```

```
Qy 1344 aggcaggccttctctgtatctctctggaagtagtggaggccacatgtgctgctgctg 1403
Db 20536 AGGCAGGCGCTTCTCTGTATCTCTGGAAGTAGTGGAGGCCACATGTGCTGCTGCTG 20595
```


	* 157686	157785:	gap of unknown length	
	* 157786	159784:	contig of 1999 bp in length	
	* 159785	159884:	gap of unknown length	
	* 159885	161347:	contig of 1463 bp in length	
	* 161348	161447:	gap of unknown length	
	* 161448	162678:	contig of 1231 bp in length	
	* 162679	162778:	gap of unknown length	
	* 162779	163996:	contig of 1218 bp in length	
	* 163997	164096:	gap of unknown length	
	* 164097	166978:	contig of 2882 bp in length	
	* 166979	167078:	gap of unknown length	
	* 167079	168859:	contig of 1781 bp in length	
	* 168860	168959:	gap of unknown length	
	* 168960	170817:	contig of 1858 bp in length	
	* 170818	170917:	gap of unknown length	
	* 170918	172018:	contig of 1101 bp in length	
	* 172019	172118:	gap of unknown length	

Query Match	7.6%;	Score 169.8;	DB 2;	Length 179953;
Best Local Similarity	85.5%;	Pred. No. 2.3e-39;		
Matches 189;	Conservative 0;	Mismatches 32;	Indels 0;	Gaps

QY	1136	aggctgctgaagaactactgtgagtgctatgagcccaaaatactgtgtcttcttcattgg	1195
Db	58966	AGCCGTGATTTCATCAAGCATTTTTTATCATCAGGCCAAATACAGTGTTCTTCCATTGG	59025
QY	1196	caaatgcattcttgcacaaaactatgaagaaagtccagacaacgaaaaatcgtgatgagcac	1255
Db	59026	CANATGCATTCTTGCAAAAACATATGAAGAAAGCCAGAGCGAAAAAATGCTGATGAGCAC	59085
QY	1256	accccactacatggagcctgggaccttggagagcagcattatttgtccccagccaagtt	1315
Db	59086	GCCCCACTACATGGAGCGCTGGGACTTTGAGAGCAGCCACCATTCTGCCAGCGCAAAT	59145
QY	1316	ctcagaacctcaaacactgacaaaaataggcaggccttct	1356
Db	59146	CTCAGGACCTCCAAGCTGAGAAAAATAGTAGGTAGGTGCTCT	59186

RESULT	7
AC097959/c	
LOCUS	
DEFINITION	AC097959 179953 bp DNA linear HTG 20-DEC-2
KEYWORDS	Rattus norvegicus clone CH230-114G10, *** SEQUENCING IN PROGRESS
SOURCE	Norway rat.
ORGANISM	Rattus norvegicus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
AUTHORS	Rattus. 1 (bases 1 to 179953)
	Murny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.B., Amaral-tunde,H.C., Are,J.R., Banks,T., Barbara,J., Benton,K.J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,L., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burckett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., Davio,R., Davila,M.L., Davis,C., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale, Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B., Homsfi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Jouhan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,

RESULT 7

AC097959/c
LOCUS

LOCUS DEFINITION

ACCESSION
VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Louisghe, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhney, E., McLeod, M. P., Meador, M.,
 Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Ogih, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojebokan, I., Rolfe, M.,
 Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
 Sisson, I., Sodergren, E., Sonaikhe, I., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telifrod, B., Thomas, N.,
 Thomas, S., Usmari, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL

2 (bases 1 to 179953)

Unpublished

Worley, K. C.

Direct Submission

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:16901822.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GFSK

Center clone name: CH230-114G10

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findPhrapList

Consensus quality: 163172 bases at least Q40

Consensus quality: 168807 bases at least Q30

Consensus quality: 174078 bases at least Q20

Estimated insert size: 160479; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 62 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 6639: contig of 6639 bp in length

* 6640 6739: gap of unknown length

* 6740 12683: contig of 5944 bp in length

* 12684 12783: gap of unknown length

* 12784 16867: contig of 4084 bp in length

* 16868 16967: gap of unknown length

* 16968 22929: contig of 5962 bp in length

* 22930 23029: gap of unknown length

* 23030 30265: contig of 7236 bp in length

* 30266 30365: gap of unknown length

* 30366 36389: contig of 6024 bp in length

* 36390 36489: gap of unknown length

* 36490 42068: contig of 5579 bp in length

* 42069 42168: gap of unknown length

* 42169 46728: contig of 4560 bp in length

* 46729 46828: gap of unknown length

* 46829 52208: contig of 5380 bp in length

* 52209 52308: gap of unknown length

* 52309 56609: contig of 4301 bp in length

* 56610 56709: gap of unknown length

* 56710 59548: contig of 2839 bp in length

* 59549 59649: gap of unknown length

* 59650 62710: contig of 3062 bp in length

* 62711 62810: gap of unknown length

* 62811 66227: contig of 3417 bp in length

* 66228 66327: gap of unknown length

* 66328 69089: contig of 2762 bp in length

* 69090 69190: gap of unknown length

* 69191 72646: contig of 3456 bp in length

* 72647 72745: gap of unknown length

* 72746 76817: contig of 4072 bp in length

* 76818 76917: gap of unknown length

* 76919 80040: contig of 3123 bp in length

* 80041 80140: gap of unknown length

* 80141 83579: contig of 3439 bp in length

* 83580 83679: gap of unknown length

* 83680 86465: contig of 2786 bp in length

* 86466 86565: gap of unknown length

* 86566 90515: contig of 3950 bp in length

* 90516 90615: gap of unknown length

* 90616 94222: contig of 3607 bp in length

* 94223 94323: gap of unknown length

* 94324 97924: contig of 3602 bp in length

* 97925 98024: gap of unknown length

* 98025 98751: contig of 1727 bp in length

* 98752 99851: gap of unknown length

* 99852 102642: contig of 2791 bp in length

* 102643 102742: gap of unknown length

* 102743 106798: contig of 4056 bp in length

* 106799 106898: gap of unknown length

* 106899 110343: contig of 3445 bp in length

* 110344 110443: gap of unknown length

* 110444 112934: contig of 2491 bp in length

* 112935 113034: gap of unknown length

* 113035 116034: contig of 3000 bp in length

* 116035 116134: gap of unknown length

* 116135 119554: contig of 3420 bp in length

* 119555 122029: contig of 2375 bp in length

* 122030 122129: gap of unknown length

* 122130 125047: contig of 2917 bp in length

* 125048 125146: gap of unknown length

* 125147 127444: contig of 2298 bp in length

* 127445 127544: gap of unknown length

* 127545 129251: contig of 1707 bp in length

* 129252 129351: gap of unknown length

* 129352 131551: contig of 2200 bp in length

* 131552 131651: gap of unknown length

* 131652 133716: contig of 2065 bp in length

* 133717 133816: gap of unknown length

* 133817 136695: contig of 2879 bp in length

* 136696 136795: gap of unknown length

* 136796 138444: contig of 1649 bp in length

* 138445 138544: gap of unknown length

* 138545 140364: contig of 1820 bp in length

* 140365 140464: gap of unknown length

* 140465 142056: contig of 1592 bp in length

* 142057 142156: gap of unknown length

* 142157 144300: contig of 2144 bp in length

* 144301 144400: gap of unknown length

* 144401 145630: contig of 1230 bp in length

* 145631 145730: gap of unknown length

* 145731 147565: contig of 1835 bp in length

* 147566 147665: gap of unknown length

* 147666 149525: contig of 1860 bp in length

* 149526 149625: gap of unknown length

* 149626 151295: contig of 1670 bp in length

* 151296 151395: gap of unknown length

* 151396 152702: contig of 1307 bp in length

* 152703 152802: gap of unknown length


```

* 152803 154292: contig of 1490 bp in length
* 154293 154392: gap of unknown length
* 154393 155990: contig of 1598 bp in length
* 155991 156090: gap of unknown length
* 156091 157685: contig of 1595 bp in length
* 157686 157785: gap of unknown length
* 157786 159784: contig of 1999 bp in length
* 159785 159884: gap of unknown length
* 159885 161347: contig of 1463 bp in length
* 161348 161447: gap of unknown length
* 161448 162678: contig of 1231 bp in length
* 162679 162778: gap of unknown length
* 162779 163996: contig of 1218 bp in length
* 163997 164096: gap of unknown length
* 164097 166978: contig of 2882 bp in length
* 166979 167078: gap of unknown length
* 167079 168859: contig of 1781 bp in length
* 168860 168959: gap of unknown length
* 168960 170817: contig of 1858 bp in length
* 170818 170917: gap of unknown length
* 170918 172018: contig of 1101 bp in length
* 172019 172118: gap of unknown length

Query Match      6.5%; Score 145.8; DB 2; Length 179953;
Best Local Similarity 92.7%; Pred. No. 4.8e-32;
Matches 153; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 495 gttgaatacaagacagtgtagtgcagcgccgagcctgaagcagcagcttcc 554
Db 115559 GTAGAATCAAGAGCAGTGCGGTGGTCCAGGAGCAGGCGCTCAAGATC 115500

QY 555 caggccctctgctcaggaatcctgttgcagctccatccagcagcagggag 614
Db 115499 CAGGCCCTCTGCTCAGGAATCCTGCTCAAGTTCATCTCATCCAGGAGGAG 115440

QY 615 gcttcagctgcctcgggaagaagactccagccatccagcagcagtgatt 659
Db 115439 GCCTCGGCTGCGCCCGGAAAAAGACTCCAGCCCATGTAAC 115395

RESULT 8
AP003096/c      157269 bp      DNA      linear      HTG 18-JAN-2001
LOCUS      Homo sapiens chromosome 11 clone CTD-2007L18 map 11q, WORKING DRAFT
DEFINITION      SEQUENCE, 32 unordered pieces.
ACCESSION      AP003096.1 GI:12275357
VERSION      HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS      Homo sapiens DNA, clone:CTD-2007L18.
SOURCE      Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 157269)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Homo sapiens 157,269 genomic DNA of 11q
JOURNAL      Published Only in DataBase (2001) In press
REFERENCE      2 (bases 1 to 157269)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE      Direct Submission
JOURNAL      Submitted (16-JAN-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp. URL:http://hgp.gsc.riken.go.jp/.
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT
----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
----- Center project name: Humdraft11

```

```

Center clone name: CTD-2007L18
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator Er-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 147566 bases at least Q40
Consensus quality: 151779 bases at least Q30
Consensus quality: 153479 bases at least Q20
Insert size: 154169; sum-of-contigs
Quality coverage: 8.32x in Q20 bases; sum-of-contigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 11499 contig of 11499 bp in length
11600 19042 contig of 7443 bp in length
19143 26427 contig of 7285 bp in length
26528 38740 contig of 12213 bp in length
38841 55483 contig of 16643 bp in length
55584 67548 contig of 11965 bp in length
67649 74728 contig of 7080 bp in length
74829 79575 contig of 4747 bp in length
79676 88957 contig of 9282 bp in length
89058 93854 contig of 4797 bp in length
93955 98914 contig of 4960 bp in length
99015 102673 contig of 3659 bp in length
102774 105297 contig of 2524 bp in length
105398 110377 contig of 4980 bp in length
110478 117540 contig of 7063 bp in length
117641 120176 contig of 2536 bp in length
120277 125707 contig of 5431 bp in length
125808 128247 contig of 2440 bp in length
128348 131115 contig of 2768 bp in length
131216 134071 contig of 2856 bp in length
134172 136288 contig of 2117 bp in length
136389 139198 contig of 2810 bp in length
139299 141579 contig of 2281 bp in length
141680 143825 contig of 2146 bp in length
143926 146191 contig of 2266 bp in length
146292 148382 contig of 2091 bp in length
148483 149666 contig of 1184 bp in length
149767 151747 contig of 1981 bp in length
151848 153751 contig of 1904 bp in length
153852 154877 contig of 1026 bp in length
154978 156158 contig of 1181 bp in length
156259 157269 contig of 1011 bp in length.

```

* NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 11499: contig of 11499 bp in length
11500 11599: gap of 100 bp
11600 19042: contig of 7443 bp in length
19043 19142: gap of 100 bp
19143 26427: contig of 7285 bp in length
26428 26527: gap of 100 bp
26528 38740: contig of 12213 bp in length
38741 38840: gap of 100 bp
38841 55483: contig of 16643 bp in length
55484 55583: gap of 100 bp
55584 67548: contig of 11965 bp in length
67549 67648: gap of 100 bp
67649 74728: contig of 7080 bp in length
74729 74828: gap of 100 bp
74829 79575: contig of 4747 bp in length

```



```

* 79576 79675: gap of 100 bp
* 79676 88957: contig of 9282 bp in length
* 88958 89057: gap of 100 bp
* 89058 93854: contig of 4797 bp in length
* 93855 93954: gap of 100 bp
* 93955 98915: contig of 4960 bp in length
* 98915 99014: gap of 100 bp
* 99015 102673: contig of 3659 bp in length
* 102674 102773: gap of 100 bp
* 102774 105297: contig of 2524 bp in length
* 105298 105398: gap of 100 bp
* 105398 110377: contig of 4980 bp in length
* 110378 110477: gap of 100 bp
* 110478 117540: contig of 7063 bp in length
* 117541 117640: gap of 100 bp
* 117641 120176: contig of 2536 bp in length
* 120177 120276: gap of 100 bp
* 120277 125707: contig of 5431 bp in length
* 125708 125807: gap of 100 bp
* 125808 128247: contig of 2440 bp in length
* 128248 128347: gap of 100 bp
* 128348 131115: contig of 2768 bp in length
* 131116 131215: gap of 100 bp
* 131216 134071: contig of 2856 bp in length
* 134072 134171: gap of 100 bp
* 134172 136288: contig of 2117 bp in length
* 136289 136388: gap of 100 bp
* 136389 139198: contig of 2810 bp in length
* 139199 139298: gap of 100 bp
* 139299 141579: contig of 2281 bp in length
* 141580 141679: gap of 100 bp
* 141680 143825: contig of 2146 bp in length
* 143826 143925: gap of 100 bp
* 143926 146191: contig of 2266 bp in length
* 146192 146291: gap of 100 bp
* 146292 148382: contig of 2091 bp in length
* 148383 148482: gap of 100 bp
* 148483 149666: contig of 1184 bp in length
* 149667 149766: gap of 100 bp
* 149767 151747: contig of 1981 bp in length
* 151748 151847: gap of 100 bp
* 151848 153751: contig of 1904 bp in length
* 153752 153851: gap of 100 bp
* 153852 154877: contig of 1026 bp in length
* 154878 154977: gap of 100 bp
* 154978 156158: contig of 1181 bp in length
* 156159 156258: gap of 100 bp
* 156259 157269: contig of 1011 bp in length.
FEATURES
    source
        1. 157269
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="11q"
            /clone="CTD-2007L18"
        1. 11499
            /note="assembly_fragment"
            11600..19042
            /note="assembly_fragment"
            19143..28427
            /note="assembly_fragment"
            26528..38740
            /note="assembly_fragment"
            38841..55483
            /note="assembly_fragment"
            55584..67548
            /note="assembly_fragment"
            67649..74728
            /note="assembly_fragment"
            74829..79575
            /note="assembly_fragment"
            79676..88957
            /note="assembly_fragment"

```

```

misc_feature 89058..93854
              /note="assembly_fragment"
misc_feature 93955..98914
              /note="assembly_fragment"
misc_feature 99015..102673
              /note="assembly_fragment"
misc_feature 102774..105297
              /note="assembly_fragment"
misc_feature 105398..110377
              /note="assembly_fragment"
misc_feature 110478..117540
              /note="assembly_fragment"
misc_feature 117641..120176
              /note="assembly_fragment"
misc_feature 120277..125707
              /note="assembly_fragment"
misc_feature 125808..128247
              /note="assembly_fragment"
misc_feature 128348..131115
              /note="assembly_fragment"
misc_feature 131216..134071
              /note="assembly_fragment"
misc_feature 134172..136288
              /note="assembly_fragment"

Query Match      6.3%; Score 141; DB 2; Length 157269;
Best Local Similarity 76.0%; Pred. No. 1.3e-30;
Matches 174; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1342 ataggcaggccttcctgtatctctctgggaagtagtgaggccacatgctgcctgcctgc 1401
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130834 ACAGCGCGCTTCTCTCATGCATCTCTCTGGAGGTGGTGGAGCCACATCGCGCTGCCCTGC 130775

Qy 1402 tggccagggtgaggaaacagacagcaggagcactgttccccaagcttgcctcagcagatga 1461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130774 TTGCTCAGGAGAGAGAGCGCCGAGAAAGAACACTGCTCCAGTGGCAGAGACATGA 130715

Qy 1462 tcctggaggaggtttggaaggtgcctgcgcagattctccacatcagttcgaagtcgaagg 1521
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130714 TCCTGGAGGAATTTGGAAGGTGCTTATCACAGATTTCTCCACACTGAGTTTAAATCTAAGG 130655

Qy 1522 ggcctgaaattgtagtagcgtgcgaagctggttaaagggggaatgcctgtgcgc 1570
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130654 GATTGAAATGGAGTAGAGATATAAGTGAATGTCATGTCATGTTGTTTGTGTC 130606

RESULT 9
AP003096      Homo sapiens chromosome 11 clone CTD-2007L18 map 11q, WORKING DRAFT
LOCUS          SEQUENCE, 32 unordered pieces.
DEFINITION     AP003096      Homo sapiens
ACCESSION      AP003096      GI:12275357
VERSION        HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS       Homo sapiens DNA, clone:CTD-2007L18.
SOURCE         Homo sapiens
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 157269)
AUTHORS        Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
               Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE          Homo sapiens 157,269 genomic DNA of 11q
JOURNAL        Published Only in DataBase (2001) In press
REFERENCE      2 (bases 1 to 157269)
AUTHORS        Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
               Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE          Direct Submission
JOURNAL        Submitted (16-JAN-2001) Masahira Hattori, The Institute of Physical
               and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
               1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
               (E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
               Tel:81-45-503-9111, Fax:81-45-503-9170)
               ----- Genome Center
               Center: RIKEN Genomic Sciences Center(GSC)
COMMENT

```

Center code: RIKEN
 Web site: <http://hgp.gsc.riken.go.jp/>
 Contact: hattori@gs.riken.go.jp

----- Project information
 Center project name: Humdraft11
 Center clone name: CTD-2007L18

----- Summary Statistics
 Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 147566 bases at least Q40
 Consensus quality: 151779 bases at least Q30
 Consensus quality: 153479 bases at least Q20
 Insert size: 154169; sum-of-contigs
 Quality coverage: 8.32x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 11499 contig of 11499 bp in length
11600 19042 contig of 7443 bp in length
19143 26427 contig of 7285 bp in length
26528 38740 contig of 12213 bp in length
38841 55483 contig of 16643 bp in length
55584 67548 contig of 11965 bp in length
67649 74728 contig of 100 bp
74829 79575 contig of 4747 bp in length
79676 88957 contig of 9282 bp in length
89058 93955 contig of 4797 bp in length
93955 99015 contig of 100 bp
99015 102673 contig of 3659 bp in length
102673 102774 contig of 100 bp
102774 105297 contig of 2524 bp in length
105297 105398 contig of 100 bp
105398 110377 contig of 4980 bp in length
110378 110477 contig of 100 bp
110478 117540 contig of 7063 bp in length
117541 117640 contig of 100 bp
117641 120176 contig of 2536 bp in length
120177 120277 contig of 100 bp
120277 125707 contig of 5431 bp in length
125708 125808 contig of 100 bp
125808 128247 contig of 2440 bp in length
128248 128347 contig of 100 bp
128348 131115 contig of 2768 bp in length
131116 131215 contig of 100 bp
131216 134071 contig of 2856 bp in length
134072 134171 contig of 100 bp
134172 136288 contig of 2117 bp in length
136289 136388 contig of 100 bp
136389 139198 contig of 2810 bp in length
139199 139298 contig of 100 bp
139299 141579 contig of 2281 bp in length
141580 141679 contig of 100 bp
141680 143825 contig of 2146 bp in length
143826 143925 contig of 100 bp
143926 146191 contig of 2266 bp in length
146192 146291 contig of 100 bp
146292 148382 contig of 2091 bp in length
148383 148482 contig of 100 bp
148483 149666 contig of 1184 bp in length
149667 149766 contig of 100 bp
149767 151747 contig of 1981 bp in length
151748 151847 contig of 100 bp
151848 153751 contig of 1904 bp in length
153752 153851 contig of 100 bp
153852 154877 contig of 1026 bp in length
154878 154977 contig of 100 bp
154978 156158 contig of 1181 bp in length
156159 156258 contig of 100 bp
156259 157269 contig of 1011 bp in length.

```

FEATURES

Source

```

1. .157269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"

```

```

1. .11499
/clone="CTD-2007L18"

```

```

misc_feature 11499
misc_feature 11600..19042
misc_feature 19143..26427
misc_feature 26528..38740
misc_feature 38841..55483
misc_feature 55584..67548
misc_feature 67649..74728

```

```

Location/Qualifiers

```

```

/misc_feature /note="assembly_fragment"
74829..79575
/misc_feature /note="assembly_fragment"
79676..88957
/misc_feature /note="assembly_fragment"
89058..93854
/misc_feature /note="assembly_fragment"
93955..98914
/misc_feature /note="assembly_fragment"
99015..102673
/misc_feature /note="assembly_fragment"
102774..105297
/misc_feature /note="assembly_fragment"
105398..110377
/misc_feature /note="assembly_fragment"
110478..117540
/misc_feature /note="assembly_fragment"
117641..120176
/misc_feature /note="assembly_fragment"
120277..125707
/misc_feature /note="assembly_fragment"
125808..128247
/misc_feature /note="assembly_fragment"
128348..131115
/misc_feature /note="assembly_fragment"
131216..134071
/misc_feature /note="assembly_fragment"
134172..136288

```

Query Match 5.5%; Score 123.6; DB 2; Length 157269;
 Best Local Similarity 72.9%; Pred. No. 2.7e-25;
 Matches 159; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```

QY 1164 tatgagggccaaatcatgtttcttcattgcaaatcattgttgcataaaactatgaa 1223
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36566 TACCAGGCCCAATATATGTCTTCTATTGCAAAATGCTGGTGCATAAATATGAA 36625
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1224 gaaatccagaaacgaaatgctgtagcagacacccactacatgagcctggggacttt 1283
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36626 GAAAGCCCAACGAAAGACATATGATGATGCAAACTACATGACAGTGGAGTTTG 36685
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1284 gagagcagcattattgtcccccagcgaagtctcagagcctccaaactgagaaaaat 1343
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36686 GAAAGCAGCCATTTACCTGCCCAACGAAATTTTCAGGACTTCCAGATTCAGTCAGCAT 36745
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1344 aggcagcctctctgtatctctcctgggaagttagga 1381
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 36746 AGGTAGGACTTTGAACCTTGAACCTTCTGCTAATAGAGGA 36783
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 10
LOCUS AP003732 188859 bp DNA linear HTG 12-JUN-2001
DEFINITION Homo sapiens chromosome 11 clone RP11-682D22 map 11q, WORKING DRAFT
SEQUENCE, 39 unordered pieces.
ACCESSION AP003732
VERSION AP003732.1 GI:14349298
KEYWORDS HTG; HTGS; PHASE1; HTGS; DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-682D22.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 188859)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (11-JUN-2001) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsr.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)

```

```

Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsr.riken.go.jp
----- Project Information
Center project name: Humdraft11
Center clone name: RP11-682D22
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 181083 bases at least Q40
Consensus quality: 183471 bases at least Q30
Consensus quality: 184447 bases at least Q20
Insert size: 185059; sum-of-contigs
Quality coverage: 7.99x in Q20 bases; sum-of-contigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 39 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 19831 contig of 19831 bp in length
19932 31497 contig of 11566 bp in length
31598 44647 contig of 13050 bp in length
44748 54198 contig of 9451 bp in length
54299 63363 contig of 9065 bp in length
63464 71678 contig of 8215 bp in length
71779 79960 contig of 8182 bp in length
80061 87061 contig of 7001 bp in length
87162 93934 contig of 6773 bp in length
94035 101099 contig of 7065 bp in length
101200 108623 contig of 7424 bp in length
108724 114505 contig of 5782 bp in length
114606 119720 contig of 5115 bp in length
119821 124688 contig of 4868 bp in length
124789 130420 contig of 5632 bp in length
130521 134923 contig of 4403 bp in length
135024 138158 contig of 3135 bp in length
138259 143086 contig of 4828 bp in length
143187 146616 contig of 3430 bp in length
146717 150135 contig of 3419 bp in length
150236 153465 contig of 3230 bp in length
153566 156263 contig of 2698 bp in length
156364 158882 contig of 2519 bp in length
158983 161109 contig of 2127 bp in length
161210 164035 contig of 2826 bp in length
164136 166612 contig of 2477 bp in length
166713 168665 contig of 1953 bp in length
168766 171614 contig of 2849 bp in length
171715 173550 contig of 1836 bp in length
173651 174290 contig of 640 bp in length
174391 176225 contig of 1835 bp in length
176326 178179 contig of 1854 bp in length
178280 179640 contig of 1361 bp in length
179741 181689 contig of 1949 bp in length
181790 183060 contig of 1271 bp in length
183161 184208 contig of 1048 bp in length
184309 185664 contig of 1356 bp in length
185765 187616 contig of 1852 bp in length
187717 188859 contig of 1143 bp in length.

```

* NOTE: This is a 'working draft' sequence. It currently consists of 39 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 19831: contig of 19831 bp in length
* 19832 19932: gap of 100 bp
* 19932 31497: contig of 11566 bp in length

```

*	31498	31597:	gap of	100 bp
*	31598	44647:	contig of 13050 bp in length	
*	44648	44747:	gap of	100 bp
*	44748	54198:	contig of 9451 bp in length	
*	54199	54298:	gap of	100 bp
*	54299	63363:	contig of 9065 bp in length	
*	63364	71678:	gap of	100 bp
*	71679	71778:	gap of	100 bp
*	71779	79960:	contig of 8182 bp in length	
*	79961	80060:	gap of	100 bp
*	80061	87061:	contig of 7001 bp in length	
*	87062	87161:	gap of	100 bp
*	87162	93934:	contig of 6773 bp in length	
*	93935	94034:	gap of	100 bp
*	94035	101095:	contig of 7065 bp in length	
*	101100	101199:	gap of	100 bp
*	101200	108623:	contig of 7424 bp in length	
*	108624	108723:	gap of	100 bp
*	108724	114505:	contig of 5782 bp in length	
*	114506	114605:	gap of	100 bp
*	114606	119720:	contig of 5115 bp in length	
*	119721	119820:	gap of	100 bp
*	119821	124688:	contig of 4868 bp in length	
*	124689	124788:	gap of	100 bp
*	124789	130420:	contig of 5632 bp in length	
*	130421	130520:	gap of	100 bp
*	130521	134923:	contig of 4403 bp in length	
*	134924	135023:	gap of	100 bp
*	135024	138159:	contig of 3135 bp in length	
*	138159	138258:	gap of	100 bp
*	138259	143086:	contig of 4828 bp in length	
*	143087	143186:	gap of	100 bp
*	143187	146616:	contig of 3430 bp in length	
*	146617	146716:	gap of	100 bp
*	146717	150135:	contig of 3419 bp in length	
*	150136	150235:	gap of	100 bp
*	150236	153465:	contig of 3230 bp in length	
*	153466	153565:	gap of	100 bp
*	153566	156263:	contig of 2698 bp in length	
*	156264	156363:	gap of	100 bp
*	156364	158882:	contig of 2519 bp in length	
*	158883	158982:	gap of	100 bp
*	158984	161109:	contig of 2127 bp in length	
*	161110	161209:	gap of	100 bp
*	161210	164035:	contig of 2826 bp in length	
*	164036	164135:	gap of	100 bp
*	164136	166612:	contig of 2477 bp in length	
*	166613	166712:	gap of	100 bp
*	166713	168665:	contig of 1953 bp in length	
*	168666	168765:	gap of	100 bp
*	168766	171614:	contig of 2849 bp in length	
*	171615	171714:	gap of	100 bp
*	171715	173550:	contig of 1836 bp in length	
*	173551	173650:	gap of	100 bp
*	173651	174290:	contig of 640 bp in length	
*	174291	174390:	gap of	100 bp
*	174391	176225:	contig of 1835 bp in length	
*	176226	176325:	gap of	100 bp
*	176326	178179:	contig of 1854 bp in length	
*	178180	178279:	gap of	100 bp
*	178280	179640:	contig of 1361 bp in length	
*	179641	179740:	gap of	100 bp
*	179741	181689:	contig of 1949 bp in length	
*	181690	181789:	gap of	100 bp
*	181790	183060:	contig of 1271 bp in length	
*	183061	183160:	gap of	100 bp
*	183161	184208:	contig of 1048 bp in length	
*	184209	184308:	gap of	100 bp
*	184309	195664:	contig of 1356 bp in length	
*	185665	185764:	gap of	100 bp
*	185765	187616:	contig of 1852 bp in length	
*	187617	187716:	gap of	100 bp

```

* 187717 188859: contig of 1143 bp in length.
FEATURES
    source                Location/Qualifiers
        1..188859
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="11"
            /map="11q"
            /clone="RP11-682D22"
            1..19831
                /note="assembly_fragment"
            19932..31497
                /note="assembly_fragment clone_end:SP6 vector_side:left"
            31598..44647
                /note="assembly_fragment"
            44748..54198
                /note="assembly_fragment"
            54299..63363
                /note="assembly_fragment"
            63464..71678
                /note="assembly_fragment"
            71779..79960
                /note="assembly_fragment"
            80061..87061
                /note="assembly_fragment"
            87162..93934
                /note="assembly_fragment"
            94035..101099
                /note="assembly_fragment"
            101200..108623
                /note="assembly_fragment"
            108724..114505
                /note="assembly_fragment"
            114606..119720

Query Match          4.8%; Score 107.4; DB:2; Length 188859;
Best Local Similarity 78.2%; Pred. No. 2.5e-20;
Matches 129; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 495 gtgaatcaaaagaagcaggtggttagtgcagcgccgcagccctgaagcagcagcttc 554
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141024 GTTGAATCAGGAGGAGGAGTGGTACTACTACAGTAAATAATCGGAGAGCAACTTTG 141083
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 555 caggccctctggctcagggaatccctgttgcaagttcccatcatccccaggaggcagaggag 614
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141084 CAGAATCTCTTGCTCAGGAATCTGTTGCAAGTTCCATCGTCCAGGAACATAGAGGAT 141143
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 615 gcctccagctgcctcctcggaagaagaagactccagccccatggtgatt 659
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141144 GCCTCCTGCTGTTCTCTTAAGAAAGATTTCCAAACCAATGTTAAATT 141188

```

RESULT	11
LOCUS	AC099385/c
DEFINITION	Rattus norvegicus chromosome Rf4 clone CH230-182G12, *** SEQUENCING IN PROGRESS ***, 90 unordered pieces.
ACCESSION	AC099385
VERSION	214055 bp DNA linear HTG 21-DEC-2001
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Norway rat. Rattus norvegicus
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 214055)
AUTHORS	Muzny,D.N., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J.J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
 Dunn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
 Duggan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Correll, J.H., Guevara, W., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
 Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
 Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, J., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Loulsegue, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
 Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
 Sisson, I., Sodergren, E., Sonaika, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmari, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczkyk, R., Wooden, S.,
 Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Unpublished
 2 (bases 1 to 214055)

Worley, K.C.

Submitted (10-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 21, 2001 this sequence version replaced gi:16901968.

COMMENT

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GKDB
 Center clone name: CH230-182G12
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to
 findPhrapList
 Consensus quality: 148220 bases at least Q40
 Consensus quality: 157150 bases at least Q30
 Consensus quality: 162432 bases at least Q20
 Estimated insert size: 127504; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 2.8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 90 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 9985: contig of 9985 bp in length
 * 9986 10085: gap of unknown length
 * 10086 15886: contig of 5801 bp in length
 * 15887 15986: gap of unknown length
 * 15987 22802: contig of 6816 bp in length

22803 22902: gap of unknown length
 22903 28825: contig of 5923 bp in length
 28826 28926: gap of unknown length
 28927 34008: contig of 5083 bp in length
 34009 34109: gap of unknown length
 34110 39103: contig of 4995 bp in length
 39104 45058: contig of 5855 bp in length
 45059 45158: gap of unknown length
 45159 49552: contig of 4394 bp in length
 49553 49652: gap of unknown length
 49653 53721: contig of 4068 bp in length
 53722 57416: contig of 3595 bp in length
 57417 61127: contig of 3612 bp in length
 61128 61227: gap of unknown length
 61228 63858: contig of 2631 bp in length
 63859 67550: gap of unknown length
 67551 70157: contig of 2507 bp in length
 70158 70257: gap of unknown length
 70258 73881: contig of 3624 bp in length
 73882 78467: contig of 4486 bp in length
 78468 81755: contig of 3188 bp in length
 81756 81856: gap of unknown length
 81857 85736: contig of 3880 bp in length
 85737 88446: gap of unknown length
 88447 90556: contig of 2011 bp in length
 90557 93742: contig of 3086 bp in length
 93743 96640: contig of 2798 bp in length
 96641 96740: gap of unknown length
 96741 98731: contig of 1991 bp in length
 98732 98831: gap of unknown length
 98832 101337: contig of 2506 bp in length
 101338 103510: contig of 2073 bp in length
 103511 103610: gap of unknown length
 103611 105626: contig of 2016 bp in length
 105627 105727: gap of unknown length
 105728 108599: contig of 2873 bp in length
 108600 110749: contig of 2050 bp in length
 110750 110849: gap of unknown length
 110850 111897: contig of 1048 bp in length
 111898 111997: gap of unknown length
 111998 114094: contig of 2097 bp in length
 114095 114194: gap of unknown length
 114195 116196: contig of 2002 bp in length
 116197 116296: gap of unknown length
 116297 118498: contig of 2202 bp in length
 118499 118598: gap of unknown length
 118599 121326: contig of 2728 bp in length
 121327 121426: gap of unknown length
 121427 123062: contig of 1636 bp in length
 123063 123162: gap of unknown length
 123163 124763: contig of 1601 bp in length
 124764 124863: gap of unknown length
 124864 126923: contig of 2060 bp in length
 126924 127023: gap of unknown length
 127024 129478: contig of 2455 bp in length
 129479 129578: gap of unknown length
 129579 131521: contig of 1943 bp in length
 131522 131621: gap of unknown length
 131622 133449: contig of 1728 bp in length
 133450 133550: gap of unknown length

```

* 133450 135573: contig of 2124 bp in length
* 135574 135673: gap of unknown length
* 135674 137246: contig of 1573 bp in length
* 137247 137346: gap of unknown length
* 137347 138640: contig of 1294 bp in length
* 138641 138740: gap of unknown length
* 138741 139772: contig of 1032 bp in length
* 139773 139872: gap of unknown length
* 139873 141176: contig of 1304 bp in length
* 141177 141278: gap of unknown length
* 141279 142378: contig of 1102 bp in length
* 142379 142478: gap of unknown length
* 142479 144070: contig of 1592 bp in length
* 144071 144170: gap of unknown length
* 144171 146690: contig of 2520 bp in length
* 146691 146790: gap of unknown length
* 146791 149890: contig of 3100 bp in length
* 149891 149990: gap of unknown length
* 149991 151499: contig of 1509 bp in length
* 151500 151599: gap of unknown length
* 151600 154256: contig of 2657 bp in length
* 154257 154356: gap of unknown length
* 154357 156559: contig of 2203 bp in length
* 156560 156659: gap of unknown length
* 156660 157925: contig of 1266 bp in length
* 157926 158025: gap of unknown length
* 158026 159291: contig of 1266 bp in length
* 159292 159391: gap of unknown length
* 159392 161143: contig of 1752 bp in length
* 161144 161243: gap of unknown length
* 161244 162894: contig of 1651 bp in length
* 162895 162994: gap of unknown length
* 162995 165164: contig of 2170 bp in length
* 165165 165264: gap of unknown length

Query Match      3.8%; Score 85.4; DB 2; Length 214055;
Best Local Similarity 71.1%; Pred. No. 1.3e-13;
Matches 113; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1024 ccataagcgctgttagatagaatctgaagctttccaaacaaatgggaagggcc 1083
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15210 CCAAAGGCGATCGCTTGCAGACACCCGGAAGCCTTCAAGCCAAAGATGGAAGAGAA 15151

QY 1084 gtctggagctgttaacttcacacagcaaggaaggtgcaactgtaagcgtcaggtgcc 1143
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15150 AAGAGGGGAGTCGGATCGCGCGCACAGCAAGGCTGTAAGTCAAGGTCAGGGTGTG 15091

QY 1144 tgaagaactctgtgagtgctatgagcgaataatcatgt 1182
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 15090 TTAATAACTACTCGAGTGTATGAGTAAGGCCCTGT 15052

```

```

RESULT 12
AC103272
LOCUS      AC103272
DEFINITION Rattus norvegicus chromosome Rf4 clone CH230-230B2, WORKING DRAFT
SEQUENCE   AC103272
ACCESSION  AC103272
VERSION    AC103272.4 GI:18201687
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 247253)
AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Anaratinge,H.C., Are,J.R., Banks,T., Barbara,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

```

```

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagó,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mahoney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S.,
Ogub,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshcarl,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 247253)
Worley,K.C.
Direct Submission
Submitted (24-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 18, 2002 this sequence version replaced gi:17974784.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GJRE
Center clone name: CH230-230B2
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 239977 bases at least Q40
Consensus quality: 241552 bases at least Q30
Consensus quality: 242554 bases at least Q20
Estimated insert size: 241759; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 5.8x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 48766: contig of 48766 bp in length
* 48767 48866: gap of unknown length

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

```

* 48867 85550: contig of 36684 bp in length
* 85551 85650: gap of unknown length
* 85651 109770: contig of 24120 bp in length
* 109771 109870: gap of unknown length
* 109871 135987: contig of 26117 bp in length
* 135988 136087: gap of unknown length
* 136088 153817: contig of 17730 bp in length
* 153818 153917: gap of unknown length
* 153918 173496: contig of 19579 bp in length
* 173497 173596: gap of unknown length
* 173597 188745: contig of 15149 bp in length
* 188746 188845: gap of unknown length
* 188846 201203: contig of 12358 bp in length
* 201204 201303: gap of unknown length
* 201304 214654: contig of 13351 bp in length
* 214655 214754: gap of unknown length
* 214755 225546: contig of 10792 bp in length
* 225547 225647: gap of unknown length
* 225648 233324: contig of 7678 bp in length
* 233325 233424: gap of unknown length
* 233425 239768: contig of 6344 bp in length
* 239769 239868: gap of unknown length
* 239869 245662: contig of 5794 bp in length
* 245663 245762: gap of unknown length
* 245763 247253: contig of 1491 bp in length.

```

FEATURES

```

Location/Qualifiers
1..247253
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/chromosome="R14"
/clone="CH230-230B2"

```

```

BASE COUNT      66265 a 56932 c 55970 g 66772 t 1314 others
ORIGIN

```

```

Query Match      3.8%; Score 85.4; DB 2; Length 247253;
Best Local Similarity 71.1%; Pred. No. 1.4e-13;
Matches 113; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```

```

QY 1024 ccataaaggcggtgttatagaaatcgtgaagctttccaaacccaaatggggaagccc 1083
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5144 CCCAAAGGCGATCGCTTGACAGAACCCGGGAAGCCCTTAAGCCAAAGATAGGAAAGGAA 5203
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1084 gctctggagctgctaaacttcacacagcaagggtgcaactgtgaagcgtcagcgtgcc 1143
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5204 AGAGGGGAGTCGATCGCGCGGCACAGCAAGGCTGTAACTGCAACAGCGTGAGGTGTC 5263
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1144 tgaagaactactgtgagtctatgagggccaaatcatgt 1182
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5264 TTTAAAACTACTGCGAGTGCTATGAGTAAGGCCCTTGT 5302
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 13

```

AC106533
LOCUS      Rattus norvegicus chromosome SA clone CH230-194C20, WORKING DRAFT
DEFINITION
SEQUENCE, 60 unordered pieces.
AC106533
AC106533.1 GI:18139057
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Norway rat.

```

ORGANISM

```

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

REFERENCE

```

AUTHORS
1 (bases 1 to 297288)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,

```

```

Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Dev-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseg,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williams,A., Wleczyk,R., Woodden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 297288)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GLGJ
Center clone name: CH230-194C20
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 270519 bases at least Q40
Consensus quality: 277312 bases at least Q30
Consensus quality: 282455 bases at least Q20
Estimated insert size: 261750; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 4.6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 64687: contig of 64687 bp in length
* 64688 64787: gap of unknown length
* 64788 88093: contig of 23306 bp in length
* 88094 88193: gap of unknown length

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

COMMENT

*	88194	110710:	contig of 22517 bp in length
*	110711	110810:	gap of unknown length
*	110811	126205:	contig of 15396 bp in length
*	126207	126306:	gap of unknown length
*	126307	140643:	contig of 14337 bp in length
*	140644	140743:	gap of unknown length
*	140744	152269:	contig of 11526 bp in length
*	152270	152369:	gap of unknown length
*	152370	161545:	contig of 9176 bp in length
*	161546	161645:	gap of unknown length
*	161646	168056:	contig of 6411 bp in length
*	168057	168156:	gap of unknown length
*	168157	174853:	contig of 6697 bp in length
*	174854	174953:	gap of unknown length
*	174954	180028:	contig of 5075 bp in length
*	180029	180128:	gap of unknown length
*	180129	185496:	contig of 5368 bp in length
*	185497	185596:	gap of unknown length
*	185597	191177:	contig of 5381 bp in length
*	191178	191277:	gap of unknown length
*	191278	195855:	contig of 4578 bp in length
*	195856	195955:	gap of unknown length
*	195956	199751:	contig of 3796 bp in length
*	199752	199851:	gap of unknown length
*	199852	203711:	contig of 3860 bp in length
*	203712	203811:	gap of unknown length
*	203812	208359:	contig of 4548 bp in length
*	208360	208459:	gap of unknown length
*	208460	213004:	contig of 4545 bp in length
*	213005	213104:	gap of unknown length
*	213105	217921:	contig of 4817 bp in length
*	217922	218021:	gap of unknown length
*	218022	220835:	contig of 2814 bp in length
*	220836	220935:	gap of unknown length
*	220936	225325:	contig of 4390 bp in length
*	225326	225425:	gap of unknown length
*	225426	227468:	contig of 2043 bp in length
*	227469	227568:	gap of unknown length
*	227569	229964:	contig of 2396 bp in length
*	229965	230064:	gap of unknown length
*	230065	233398:	contig of 3334 bp in length
*	233399	233498:	gap of unknown length
*	233499	235100:	contig of 1602 bp in length
*	235101	235200:	gap of unknown length
*	235201	237615:	contig of 2415 bp in length
*	237616	237715:	gap of unknown length
*	237716	240281:	contig of 2566 bp in length
*	240282	240381:	gap of unknown length
*	240382	242648:	contig of 2267 bp in length
*	242649	242748:	gap of unknown length
*	242749	244528:	contig of 1780 bp in length
*	244529	244628:	gap of unknown length
*	244629	246520:	contig of 1892 bp in length
*	246521	246620:	gap of unknown length
*	246621	248761:	contig of 2141 bp in length
*	248762	248861:	gap of unknown length
*	248862	250054:	contig of 1193 bp in length
*	250055	250154:	gap of unknown length
*	250155	252078:	contig of 1924 bp in length
*	252079	252178:	gap of unknown length
*	252179	253921:	contig of 1743 bp in length
*	253922	254021:	gap of unknown length
*	254022	256942:	contig of 2921 bp in length
*	256943	257042:	gap of unknown length
*	257043	258216:	contig of 1174 bp in length
*	258217	258316:	gap of unknown length
*	258317	259807:	contig of 1491 bp in length
*	259808	259907:	gap of unknown length
*	259908	261229:	contig of 1322 bp in length
*	261230	261329:	gap of unknown length
*	261330	263581:	contig of 2252 bp in length
*	263582	263681:	gap of unknown length
*	263682	264918:	contig of 1337 bp in length

*	264919	265018:	gap of	unknown	length
*	265019	267542:	contig of	2524	bp in length
*	267543	267642:	gap of	unknown	length
*	267643	268890:	contig of	1248	bp in length
*	268891	268990:	gap of	unknown	length
*	268991	270550:	contig of	1560	bp in length
*	270551	270650:	gap of	unknown	length
*	270651	271789:	contig of	1139	bp in length
*	271790	271889:	gap of	unknown	length
*	271890	273613:	contig of	1724	bp in length
*	273614	273713:	gap of	unknown	length
*	273714	275207:	contig of	1494	bp in length
*	275208	275307:	gap of	unknown	length
*	275308	276330:	contig of	1023	bp in length
*	276331	276430:	gap of	unknown	length
*	276431	277994:	contig of	1564	bp in length
*	277995	278094:	gap of	unknown	length
*	278095	279151:	contig of	1057	bp in length
*	279152	279251:	gap of	unknown	length
*	279252	281226:	contig of	1975	bp in length
*	281227	281326:	gap of	unknown	length
*	281327	282391:	contig of	1065	bp in length
*	282392	282491:	gap of	unknown	length
*	282492	284153:	contig of	1662	bp in length
*	284154	284253:	gap of	unknown	length
*	284254	285853:	contig of	1600	bp in length
*	285854	285953:	gap of	unknown	length
*	285954	287690:	contig of	1737	bp in length
*	287691	287790:	gap of	unknown	length
*	287791	289023:	contig of	1233	bp in length
*	289024	289123:	gap of	unknown	length
*	289124	290148:	contig of	1025	bp in length
*	290149	290248:	gap of	unknown	length
*	290249	291618:	contig of	1370	bp in length
*	291619	291718:	gap of	unknown	length

Query Match	Score	DB 2;	Length
3.88;	Score 85.4;	DB 2;	Length 297288;

Query Match	3.6%	Score 63.4, DB 2, length 297266,
Best Local Similarity	71.1%	Pred. No. 1.4e-13;
Matches 113: Conservative	0;	Mismatches 46; Indels 0; Gaps 0;

Qy 1024 ccataaaqqcqtqtcttqatagaaatcctqaaqctttccaacccaaaaatqqgqaaaqqcc 1083

Db 284797 CCCAAAAGGCATGCCCTTGACAGAAACCCGGAGCCCTTCAAGCCCAAGATAGCAAAAGGAA 284856

Oy 1084 ntctggagctactaaacttcgacacagcaaaaggtgcaactgtaaaggctcaggctgcc 1143

DB 284857 AACAGCGCGGAGTCCGCATCCGCGCGGCACACGCACACGGCTGTAACTGCACAAACGGTCAAGGGTGTC 284916

1144 tgaagaaactactctatgaatgcctatgaagccaaaaatcatgt 1182

[illegible]

RESULT 14

AC021105	LOCUS	AC021105	187360 bp	DNA	linear	PRI 07-NOV-2001
	DEFINITION	Homo sapiens BAC clone RP11-163O17			from 4,	complete sequence.

ACCESSION AC021105

VERSION	AC0211
VERSION	110

KEYWORDS
SOURCE

SOURCE ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE	1 (bases 1 to 187360)
1	(bases 1 to 187360)

AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence

TITLE
JOURNAL
Genome Res. 8 {11}, 1097-1108 (1998)

MEDLINE 99063792

REFERENCE 2 (bases 1 to 187360)

AUTHORS
Sun, H., Abbott, A. and Spalding, L.

TITLE The sequence of unpublished

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 187360)
Waterston,R.H.
Direct Submission
Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 187360)
Waterston,R.H.
Direct Submission
Submitted (03-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 187360)
Waterston,R.
Direct Submission
Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 3, 2001 this sequence version replaced gi:13431110.

REFERENCE
AUTHORS
TITLE
JOURNAL

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics

Center project name: H_NH0163017

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanesi,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-163017;
Actual end is at base position 187360 of RP11-163017.

The region between bases 32935-33083 is single stranded and the sequence fidelity can not be guaranteed. RP11-163017 contains an unresolved dinucleotide repeat between bases 47950-48140 where the sequence fidelity can not be guaranteed. Digest information suggests that approximately 88 bases are missing from the current assembly.

FEATURES
source

Location/Qualifiers
1..187360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-163017"

```
/clone_lib="RPCI-11"
1..85      misc_feature
/note="similar to EST BE614739 (NID:g9896336)"
1..85      misc_feature
/note="similar to EST BE786771 (NID:g10207969)"
1..85      misc_feature
/note="similar to EST AV610323 (NID:g9745993)"
1..85      misc_feature
/note="similar to EST AU120774 (NID:g10936009)"
1..85      misc_feature
/note="similar to EST A1178109 (NID:g3728747)"
1..85      misc_feature
/note="similar to EST BE074520 (NID:g8422511)"
1..85      misc_feature
/note="similar to EST AA397735 (NID:g2050626) zt86a05.r1"
209..800   misc_feature
/note="similar to EST BG432741 (NID:g13339247)"
227..725   misc_feature
/note="similar to EST BF911522 (NID:g12302980)"
649..800   misc_feature
/note="similar to EST A1178109 (NID:g3728747)"
649..778   misc_feature
/note="similar to EST BE786771 (NID:g10207969)"
649..800   misc_feature
/note="similar to EST AU120774 (NID:g10936009)"
649..800   misc_feature
/note="similar to EST AV610323 (NID:g9745993)"
649..800   misc_feature
/note="similar to EST AA397735 (NID:g2050626) zt86a05.r1"
649..800   misc_feature
/note="similar to EST BE614739 (NID:g9896336)"
649..777   misc_feature
/note="similar to EST BE074520 (NID:g8422511)"
651..800   misc_feature
/note="similar to EST BG703239 (NID:g13975373)"
655..800   misc_feature
/note="similar to EST T32984 (NID:g615082)"
1023..1310 repeat_region
/rpt_family="Alu"
1309..1334 repeat_region
/rpt_family="(TTAA)n"
1678..1878 repeat_region
/rpt_family="MIR"
1938..1965 repeat_region
/rpt_family="AT-rich"
2379..2416 repeat_region
/rpt_family="AT-rich"
2399..2711 repeat_region
/rpt_family="Alu"
3034..3091 repeat_region
/rpt_family="MIR"
3295..3779 repeat_region
/rpt_family="L1"
3784..3834 repeat_region
/rpt_family="L1"
3913..4049 repeat_region
/rpt_family="MIR"
4261..4318 misc_feature
/note="similar to EST AA397735 (NID:g2050626) zt86a05.r1"
4261..4340 misc_feature
/note="similar to EST BG432741 (NID:g13339247)"
4261..4340 misc_feature
/note="similar to EST AU120774 (NID:g10936009)"
4261..4340 misc_feature
/note="similar to EST BE614739 (NID:g9896336)"
4261..4340 misc_feature
/note="similar to EST A1178109 (NID:g3728747)"
4261..4340 misc_feature
/note="similar to EST BG703239 (NID:g13975373)"
4261..4340 misc_feature
/note="similar to EST T32984 (NID:g615082)"
4261..4340 misc_feature
/note="similar to EST AV610323 (NID:g9745993)"
```

COMMENT
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by: Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jspiehl@watson.wustl.edu

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

FEATURES	Location/Qualifiers
SOURCE	1. 40893
	/organism="Caenorhabditis b
	/strain="GutAraT G16"
	/db_xref="taxon:6238"
	/clone="G45E19"
BASE COUNT	13086 a 7791 c 7927 g 12089 t
ORIGIN	

Query Match	3.4%	Score 76.4	DB 3	Length 40893
Best Local Similarity	58.4%	Pred. NO. 4.8e-11		
Matches 181	Conservative 0	Mismatches 111	Indels 18	Gaps 2
Qy 932	tcgtctcgtggtacgtgactctctccagcggggactctcgaacagctcgcagctg---	988		
Db 37990	TCTGAACCTCTACTGTGACTGCTTCGCGAACGGAGAGTTTTGCCGTGACTGCACTGC	38049		
Qy 989	-----caacacacctgcgcctagctcgagcgcttcaaaagccataaaagcgctgtct	1039		
Db 38050	GGACTGCCACAAACAATATAGATTATGACAGCCACGATCAAAAGCCATTCGTCAGTC	38109		
Qy 1040	tgatagaaatcctgaagctttccaccacaaaatgggaaagccctctgggagctgctaa	1099		
Db 38110	GGAGCGAAATCCAAATGCTTTCAAGCCGAAAATCGAATAGCTCGTGGTGGAAACCGCG	38169		
Qy 1100	acttcgacacagc-----aaagggtgcaactgtaagcgctcaggctgcctgaaagaacta	1153		
Db 38170	TATTGAACGCTTGCAATCAAAAGGGATGCCATCGCAAGAAAACGCGTTCCTGAAAAAC	38229		
Qy 1154	ctgtgagtgcctatgagcccaaaatcatgtgtcttcccatttgcaaatgcattgcttgcaa	1213		
Db 38230	TTGGAGTGTCTACGAAGCAAAAGTGCCCTGTACTGATCGATGCAAGTCAAAAGGCTGT	38289		
Qy 1214	aaactatgaa 1223			
Db 38290	GAATACGGAA 38299			

Search completed: July 11, 2002, 11:10:47
Job time: 10564 sec

[illegible]

RESULT 15
CBRG45E19 linear INV 04-NOV-2000
LOCUS CBRG45E19 40893 bp DNA
DEFINITION Caenorhabditis briggsae cosmid G45E19, complete sequence.
ACCESSION AC084631
VERSION AC084631.1 GI:11095080
KEYWORDS HTG.
SOURCE Caenorhabditis briggsae.
ORGANISM Caenorhabditis briggsae
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 40893)
AUTHORS Washington University Genome Sequencing Center.
TITLE The C. briggsae Genome Sequencing Project
JOURNAL unpublished
REFERENCE 2 (bases 1 to 40893)
AUTHORS waterston,R.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2000) Department of Genetics, Washington

Caenorhabditis ele
Drosophila melanog
Drosophila melanog

Human polynucleoti

Human OREFX OREF1979
Human colon cancer

Human colon cancer
cDNA from clone ga
Coding sequence fo

Drosophila melanog

Human breast cancer
Apoptin protein AT

Human polynucleoti
Human secreted proHuman secreted pro
Human colon cancerHuman breast cancer
Human colon cancerHuman breast and o
sequence of cloneHuman ICAM-1. Hom
Intracellular adhe

Human TCAM-1 cDNA

Human ICAM-1 (HL-60 Intracellular adhe

HL-cDNA clone encodes

Human colon cancer

Human prostate cancer

D NO:1.
spermatocyte;
atory actor;

QY 1861 ttattatgagccagcctcaagttagagctaggtctcttctcaggtggaactctgcccac 1920
 Db 1861 ttattatgagccagcctcaagttagagctaggtctcttctcaggtggaactctgcccac 1920
 QY 1921 atcacatacaagtcagggtgagccatcagggtgttttccagggccagcctgtgacaggagat 1980
 Db 1921 atcacatacaagtcagggtgagccatcagggtgttttccagggccagcctgtgacaggagat 1980
 QY 1981 atgggaggggggtcggttagagctgggtttgtttggatttttttctgttttttcttct 2040
 Db 1981 atgggaggggggtcggttagagctgggtttgtttggatttttttctgttttttcttct 2040
 QY 2041 gtatttctgcttgagtgagaaaactgtctctgtccaaacttttccataaattactg 2100
 Db 2041 gtatttctgcttgagtgagaaaactgtctctgtccaaacttttccataaattactg 2100
 QY 2101 ctgacaggtcgctgctgacacagtcacagtcacacagtcacacacaggaagtggtggtt 2160
 Db 2101 ctgacaggtcgctgctgacacagtcacagtcacacagtcacacacaggaagtggtggtt 2160
 QY 2161 attatgccacactgtgtttgtttgttgagataaaactttccagactcccaaaaaaa 2220
 Db 2161 attatgccacactgtgtttgtttgttgagataaaactttccagactcccaaaaaaa 2220
 QY 2221 aaaaaaaaaaaaaaaaaa 2241
 Db 2221 aaaaaaaaaaaaaaaaaa 2241

RESULT 2

AAZ88157
 ID AAZ88157 standard; cDNA; 1861 BP.

XX AC AAZ88157;

XX XX 25-APR-2000 (first entry)
 XX DE Mouse testis specific factor tesmin encoding cDNA SEQ ID NO:2.

XX KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 XX KW differentiation regulatory factor; male germ cell regulatory actor;
 XX KW germ cell differentiation; sterility; ss.

XX OS Mus musculus.

XX FH Key Location/Qualifiers
 XX FT CDS 271..1158
 XX FT /*tag= a
 XX FT /product= "tesmin"

XX PN WO200004147-A1.

XX XX 27-JAN-2000.

XX PF 16-JUL-1999; 99WO-JP03859.

XX PR 17-JUL-1998; 98JP-0219856.

XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

XX DR WPI; 2000-147785/13.

XX DR P-PSDB; AAY68463.

XX PT New male germ cell regulatory factor tesmin expressed in spermatocytes
 XX PT useful for investigation of germ cell differentiation and sterility -

XX PS Example 2; Page 38-42; 63pp; Japanese.

XX XX The present sequence encodes a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin. Tesmin

CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.

XX Sequence 1861 BP; 456 A; 481 C; 498 G; 426 T; 0 other;

Query Match 78.0%; Score 1747; DB 21; Length 1861;
 Best Local Similarity 100.0%; 'Pred. No. 0;
 Matches 1747; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 gttgaatacaaaagacagcaggtgttagtgccagcggcagccctgaagacagcagcttcc 554
 Db 115 gttgaatacaaaagacagcaggtgttagtgccagcggcagccctgaagacagcagcttcc 174
 QY 555 caggccctctggtcaggaaatcctgttgcaagtcccatccatccagagcagaggag 614
 Db 175 caggccctctggtcaggaaatcctgttgcaagtcccatccatccagagcagaggag 234
 QY 615 gctccagctgcctcggaaagaaactccagcccatgtgtgatgttcagctgaaagga 674
 Db 235 gctccagctgcctcggaaagaaactccagcccatgtgtgatgttcagctgaaagga 294
 QY 675 ggcgccagatgctctgcatagacaactgtgctgcgagggagctcaaaagcgtccatctg 734
 Db 295 ggcgccagatgctctgcatagacaactgtgctgcgagggagctcaaaagcgtccatctg 354
 QY 735 ctctctcagtcagatgaccagagcaggtttccctcagtcagagcagcctcaagcacaatgaca 794
 Db 355 ctctctcagtcagatgaccagagcaggtttccctcagtcagagcagcctcaagcacaatgaca 414
 QY 795 actttagtgggaagactctccagtcaccagcgaagttaaatctcatcacacaggttgtat 854
 Db 415 actttagtgggaagactctccagtcaccagcgaagttaaatctcatcacacaggttgtat 474
 QY 855 aatggagctctcccatcagctgcaatggggtgccttccctctcagcctgacctgtcga 914
 Db 475 aatggagctctcccatcagctgcaatggggtgccttccctctcagcctgacctgtcga 534
 QY 915 gggccaccccaataaactctgtctgggtactgtgactgttctccagcgggactctctgc 974
 Db 535 gggccaccccaataaactctgtctgggtactgtgactgttctccagcgggactctctgc 594
 QY 975 aacagctgcagctgcaacaactcgcgcctgagctgcagcgttcaaaagcctaagaggcg 1034
 Db 595 aacagctgcagctgcaacaactcgcgcctgagctgcagcgttcaaaagcctaagaggcg 654
 QY 1035 tgtcttgatagaaatcctgaaagcttcccaacaaaatgggaaagcgcgtctggagct 1094
 Db 655 tgtcttgatagaaatcctgaaagcttcccaacaaaatgggaaagcgcgtctggagct 714
 QY 1095 gctaaacttcgacacagcagcaaaagggtgcaactgtaagcgtcaggtgcctgaagaactac 1154
 Db 715 gctaaacttcgacacagcagcaaaagggtgcaactgtaagcgtcaggtgcctgaagaactac 774
 QY 1155 tgtgagtgctatgagcgaataatcatgtgttctccatttgcaaaatgcattgtctgcaaa 1214
 Db 775 tgtgagtgctatgagcgaataatcatgtgttctccatttgcaaaatgcattgtctgcaaa 834
 QY 1215 aactatgaagaagtcacgaacgaataatgctgatgacacaccccccactacatggagcct 1274
 Db 835 aactatgaagaagtcacgaacgaataatgctgatgacacaccccccactacatggagcct 894
 QY 1275 ggggactttgagagcagcagcattattgtcccccagcaaatctcagagcactcccaaaactg 1334
 Db 895 ggggactttgagagcagcagcattattgtcccccagcaaatctcagagcactcccaaaactg 954
 QY 1335 agaaaaaatagcagcagccttctcctgtatctctgggaagtagtgagggccacatgtgcc 1394
 Db 955 agaaaaaatagcagcagccttctcctgtatctctgggaagtagtgagggccacatgtgcc 1014
 QY 1395 tgcctgtgctgcccaggtgaggaagcagagcagggagcactgttccccagacttgctgag 1454
 Db 1015 tgcctgtgctgcccaggtgaggaagcagagcagggagcactgttccccagacttgctgag 1074

Qy 1455 cagatgacccctggaggagtttggaaggtgcctgtcgcagattctccacatcagattcaag 1514
 |||||
 Db 1075 cagatgacccctggaggagtttggaaggtgcctgtcgcagattctccacatcagattcaag 1134
 |||||
 Qy 1515 tccaagggtgaaaaattgagtagctgcaagctggtaaaagggaatgcctgtggaagc 1574
 |||||
 Db 1135 tccaagggtgaaaaattgagtagctgcaagctggtaaaagggaatgcctgtggaagc 1194
 |||||
 Qy 1575 ctacgacctgggaatctgcaccaggaagctggtgccaggaggagagcagagggccgcga 1634
 |||||
 Db 1195 ctacgacctgggaatctgcaccaggaagctggtgccaggaggagagcagagggccgcga 1254
 |||||
 Qy 1635 tcatggcaggtcagctgtgaggtctgagtgatctgcatggtactggtccagcctactcaa 1694
 |||||
 Db 1255 tcatggcaggtcagctgtgaggtctgagtgatctgcatggtactggtccagcctactcaa 1314
 |||||
 Qy 1695 ggtatctaaagtcaagcaggcgagccaccctggggatggagacactggccctcctgtcc 1754
 |||||
 Db 1315 ggtatctaaagtcaagcaggcgagccaccctggggatggagacactggccctcctgtcc 1374
 |||||
 Qy 1755 ctggggagccctctgggacctccctgcctgcataaaagaggtgattttctactgt 1814
 |||||
 Db 1375 ctggggagccctctgggacctccctgcctgcataaaagaggtgattttctactgt 1434
 |||||
 Qy 1815 tgtatgtgttcttcaaaattgcttagtagtacctccctcattcaagtatttatgagccag 1874
 |||||
 Db 1435 tgtatgtgttcttcaaaattgcttagtagtacctccctcattcaagtatttatgagccag 1494
 |||||
 Qy 1875 cctcaagttagagcagtaggctctctctcaggtggaactgtgcccaaatcacatacaagtc 1934
 |||||
 Db 1495 cctcaagttagagcagtaggctctctctcaggtggaactgtgcccaaatcacatacaagtc 1554
 |||||
 Qy 1935 agtggccatcaggggtttttccaggcagcctgtgacaggagatatggagggggtc 1994
 |||||
 Db 1555 agtggccatcaggggtttttccaggcagcctgtgacaggagatatggagggggtc 1614
 |||||
 Qy 1995 ggggttagagctgggtttgttggatttttgcgtttttttctctctgtatttctgttga 2054
 |||||
 Db 1615 ggggttagagctgggtttgttggatttttgcgtttttttctctctgtatttctgttga 1674
 |||||
 Qy 2055 agtgagaaaaactgtctctgtccaaactttctccataattactgtgcaggtgcct 2114
 |||||
 Db 1675 agtgagaaaaactgtctctgtccaaactttctccataattactgtgcaggtgcct 1734
 |||||
 Qy 2115 gctgaccagtcacagtgacctcagacaccagagaggtgaggtgtattatgcccacact 2174
 |||||
 Db 1735 gctgaccagtcacagtgacctcagacaccagagaggtgaggtgtattatgcccacact 1794
 |||||
 Qy 2175 ttgtgtttgtgtgagataaaactttccagactcccaaaaaaaataaaataaaataaa 2234
 |||||
 Db 1795 ttgtgtttgtgtgagataaaactttccagactcccaaaaaaaataaaataaaataaa 1854
 |||||
 Qy 2235 aaaaaa 2241
 |||||
 Db 1855 aaaaaa 1861
 |||||

RESULT 3

AAZ88156
 ID AAZ88156 standard; cDNA; 2134 BP.

XX
 AC AAZ88156;

XX
 DT 25-APR-2000 (first entry)

XX
 DE Human testis specific factor tesmin encoding cDNA SEQ ID NO:3.

XX
 KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 KW differentiation regulatory factor; male germ cell regulatory actor;
 KW germ cell differentiation; sterility; ss.

XX
 OS Homo sapiens.

XX Key Location/Qualifiers
 FH 407..1306
 FT /*tag= a
 FT /product= "tesmin"
 XX WO200004147-A1.
 PN 27-JAN-2000.
 PD
 XX 16-JUL-1999; 99WO-JP03859.
 PF
 XX 17-JUL-1998; 98JP-0219856.
 PR
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 XX WPI; 2000-147785/13.
 XX P-PSDB; AAY68464.
 XX
 PT New male germ cell regulatory factor tesmin expressed in spermatocytes
 PT useful for investigation of germ cell differentiation and sterility -
 XX Claim 3; Page 42-47; 63pp; Japanese.
 PS
 XX The present sequence encodes a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin. Tesmin
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX
 SQ Sequence 2134 BP; 586 A; 490 C; 528 G; 530 T; 0 other;

Query Match 28.9%; Score 647.8; DB 21; Length 2134;

Best Local Similarity 76.1%; Pred. No. 1.3e-169;

Matches 829; Conservative 0; Mismatches 247; Indels 13; Gaps 2;

Qy 495 gtgaaatacaag-aacagatgtagtgccaggcggcagccctgaagacgagcttt 553
 |||||
 Db 250 gttgaaatacaaggcaagcaggtggctactactacaagtaataatccggaagaagcaactt 309
 |||||
 Qy 554 ccaggccctctgctcaggaatcctgttcaagttcccatcattccagaggcagagga 613
 |||||
 Db 310 gcagaattcttctgctcaggaatcctgttcaagttcccatggtccaggaactagagga 369
 |||||
 Qy 614 ggcctcagctgcctcgggaagaagactccagcccatggtgattgtcagctgaaag 673
 |||||
 Db 370 tgcctcctgctgttctcttaagaagaattccaaacccaatggtgatgccaattgaaag 429
 |||||
 Qy 674 aggcgccagatgctctgcatagacaactgtggcggagggagctcaagcgctccatct 733
 |||||
 Db 430 gggcacacaatgctatgtagacaattcttagaagaagagactaaagacactccattt 489
 |||||
 Qy 734 gcttcctcagtagatgaccagagcagtttccctcagtcagagctccctaaagcactgac 793
 |||||
 Db 490 ggttcctcagtagatgaccagagcagtttccctcagtcagagctccctaaagcactgac 549
 |||||
 Qy 794 aactttagtgggaagacttctgacagtagcaggaagttaaatctcatcacacaggttga 853
 |||||
 Db 550 tgccttagtagggagatttttgcagcatcaacaaaataataatctctacacacact 609
 |||||
 Qy 854 taatgagactctccatcagctgcaatgggctgcttccctcctctgacctgcttga 913
 |||||
 Db 610 tggggggaccttaccatcggtagccaacgggtcttctccctcgggactcaactctcc 669
 |||||
 Qy 914 agggcccccacaaaataactctgtggtgactgtgactgtcttccagcggggactctg 973
 |||||
 Db 670 aggaccacaaaataacttggctggtagctgtgactgtcttccagtaggggactttg 729
 |||||
 Qy 974 caacagctgca-----gctgcaacaacactcgcccatgagctcagagcttcaa 1021
 |||||

Db 730 caacaactgcaattgtaataattgttgcacaactgtcatcatgatattgaacggttttaa 789
Qy 1022 agccataaagcgctgtctgtatagaaaacctgaagctttccaaaccaaataatgggaaagg 1081
Db 790 agccattaaaggcatgtcttggcagaaaaccagaagctttccagcccaaaaatgggaaagg 849
Qy 1082 ccgtctggagctgtctaaacttcgacacagcaaaagggtgcaactgtaagcgctcaggctg 1141
Db 850 ccaattgggcaatgtcaagcccgacacaaaggggtgcaactgcaggaggtcaggctg 909
Qy 1142 cctgaagaactactgtagctgactgaagccaaataatctgttctccatttgcaaatg 1201
Db 910 cctgaagaataactgcgagtgctatgagggcccaaatattgttcttctatttgcaaatg 969
Qy 1202 cattcttgcaaaaactatgaagaagtcacagacagaaaggtgctgatgagcacacccca 1261
Db 970 cattgttgcaaaaattatgaagaagcccgagacgaaagacactcaatgagcatgccaaa 1029
Qy 1262 ctacatgagcctgggactttgagagcagccattattgtccccagccaaagtcttcagg 1321
Db 1030 ctacatgcagactggaggtttggaaggcagccattacctgccacacaaagaaatttcagg 1089
Qy 1322 acctccaaaactgaaaaataagcagggcctctctctgtatctctctgggaagtgtgga 1381
Db 1090 acttccaaagattcagtcacgatagcggcctctctcatgcatctcctggaggtgtgga 1149
Qy 1382 ggcacatgtcctgcctgcctgcccaggggtgagaagcagagcaggagcactgttcccc 1441
Db 1150 ggcacacagcgcctgcctgctcaggggaggaagcagggcgaagacactgctccaa 1209
Qy 1442 aagcttgctgagcagatgactcctgagaggtttggaaggtgcctgtcgcagattctcca 1501
Db 1210 gtgcctggcagcagcagatgactcctggagaatttggaaggtgcttatcacagattctcca 1269
Qy 1502 catcagattcaagtcacaggggctgaaatgtgagtagcgtgcaagctggttaagggggaat 1561
Db 1270 cactgagtttaactaaagattgaaagtgaaatggagtagagataaaagtgtgaatgatgtt 1329
Qy 1562 gctgtgtgc 1570
Db 1330 gatattgtc 1338

RESULT 4
ID AAL01300 standard; cDNA; 445 BP.
XX
AC AAL01300;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen cDNA SEQ ID NO: 1301.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ss.
OS Homo sapiens.
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.

[illegible]

```
XX 28-FEB-2000; 2000US-0515126.
PR 18-MAR-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT;
XX WPI: 2001-514838/56.
XX P-PSDB; AAO09157.
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders
XX Claim 1; SEQ ID NO 9148; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 387 BP; 127 A; 65 C; 63 G; 119 T; 13 other;

Query Match      2.0%; Score 45.2; DB 22; Length 387;
Best Local Similarity 66.3%; Pred. No. 0.03;
Matches 65; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 2144 agaagtgaggtgcttatatgccacattgtgtgtgtgtgtgtgagataaacctttc 2203
DB 110 agttgttagaagatgggtattgtcactttttttttgtgaaaaaaccttcc 169

QY 2204 cagactcccaaaaaaaaaaaaaaaaaaaaaaaaaa 2241
DB 170 cttccaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 207

RESULT 15
AAF16214
ID AAF16214 standard; cDNA; 1696 BP.
XX AAF16214;
AC AAF16214;
XX 13-MAR-2001 (first entry)
XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:649.
XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX Homo sapiens.
OS
XX WO200055174-A1.
PN
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05988.
XX
```

```
PR 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX Rosen CA, Ruben SM;
XX WPI: 2000-587513/55.
XX P-PSDB; AAB57011.
XX Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer
XX Claim 1; Page 1090; 2338pp; English.
XX AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX Sequence 1696 BP; 452 A; 465 C; 430 G; 347 T; 2 other;

Query Match      2.0%; Score 44.8; DB 21; Length 1696;
Best Local Similarity 81.2%; Pred. No. 0.087;
Matches 52; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 2178 tgtttgtgtgagaataaacctttccagactcccaaaaaaaaaaaaaaaaaaaaaa 2237
DB 1611 tgttttaaatgacataaaacagccagaatgtcaaaaaaaaaaaaaaaaaaaaaa 1670

QY 2238 aaaa 2241
DB 1671 aaaa 1674
```

Search completed: July 11, 2002, 10:54:07
Job time: 9564 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 08:14:43 ; Search time 3455.78 Seconds
(without alignments)
8752.489 Million cell updates/sec

Title: US-09-743-237-1
Perfect score: 2241
Sequence: 1 tatcctgtggttgcccg.....aaaaaaaaaaaaaaaaaaaaa 2241

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1291.4	57.6	1730	11 AK015724	Mus muscu
c	647.2	28.9	708	10 BG076362	H3158G12-
3	604.8	27.0	631	9 AV282581	AV282581
4	505.8	22.6	544	10 BG101151	uy20a06.y
5	499.2	22.3	545	12 AZ661404	IM0540B04
6	493.2	22.0	505	10 BG088886	H3158G12-
7	449.0	20.0	2010	11 AK015732	Mus muscu
c	440.2	19.6	468	9 AW558944	AW558944
9	379.8	16.9	418	10 BF452508	uz27g12.y
10	341.8	15.3	842	10 B1829802	603079876
c	333	14.9	374	10 BF152725	uz95a04.x
12	314.8	14.0	329	12 AZ660217	IM0538L08
13	265.4	11.9	768	10 BG977029	602845588
14	227.4	10.1	276	9 B328330	BB328330
15	224.4	10.0	247	9 AV254566	AV254566
16	221	9.9	275	9 AV260483	AV260483
17	202.4	9.0	654	9 BB617220	BB617220

18	194.6	8.7	244	9	AV256079	AV256079
19	184.6	8.2	241	9	AV045606	AV045606
20	175.6	7.8	283	9	BB068350	BB068350
21	173	7.7	697	9	BB611478	BB611478
22	171.8	7.7	312	9	AV277921	AV277921
23	171.4	7.6	574	9	AI981460	pat.pk005
24	171	7.6	333	9	AL040101	DKF2P434C
25	167.4	7.5	221	9	BB536930	BB536930
26	166.8	7.4	257	9	AV265406	AV265406
c	166	7.4	420	10	BG272409	nan30a05.
28	164.2	7.3	475	10	BI706999	fq13f03.y
29	159.8	7.1	266	9	AF012383	AF012383
30	149.4	6.7	397	10	BF937310	fm3c07.y
31	149	6.6	704	9	AV715308	AV715308
32	145.6	6.5	176	9	AV261053	AV261053
33	138.8	6.2	204	9	AV343095	AV343095
34	133.8	6.0	241	9	AV282159	AV282159
35	133.6	6.0	638	10	BF426275	df69g02.y
36	124.8	5.6	285	9	AV091958	AV091958
c	123	5.5	696	12	AZ512200	IM0357P04
38	120	5.4	312	9	AW270088	xv37g12.x
39	111.6	5.0	118	10	BF152726	uz95a04.y
c	107.4	4.8	442	9	AI307797	tb28c05.x
41	103	4.6	918	10	BG916417	602814015
42	95.2	4.2	888	10	BG773009	602721268
c	94	4.2	642	9	AI968415	wu02b05.x
44	86.6	3.9	271	10	BG184200	RST3121.A
45	74.4	3.3	552	10	BI421583	EST532249

ALIGNMENTS

RESULT 1

AK015724	AK015724	1730 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930507A18:metallothionein-like 5, testis-specific (tesmin), full insert sequence.				
DEFINITION					
ACCESSION	AK015724				
VERSION	AK015724.1	GI:12854168			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	Carninci,P. and Hayashizaki,Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2 (sites)				
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	204999374				
PUBMED	11042159				
REFERENCE	3 (sites)				
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.				
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format				
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)				

Qy	1694	aggatcctaagaatgcgaagcagccacctggtgggattgagacattgcccctctctgtc	1753
Db	361	AGGTATCTCTAAAGTGCAGGACGAGAGCCACCTGTGGATGTGACACTGGGCCCTCTCTGTCT	420
Qy	1754	cttgggagggccctctggggaactccctgcccctgcataaaaagaggggtgattttctactctg	1813
Db	421	CCTGGGAGGCGCCTCTCTGGGACTCCCTGCGCTGCATAAAGAGGGTGATTTTCTACTTG	480
Qy	1814	ttgttatgtttgtcttccaattgcttagtagtacctccattccaagtattattgagcca	1873
Db	481	TTGTTATGTGTTTCTCTTCAAAATTGCTTAGTAGTACCTCCATTCAAGTTATTATGAGCCA	540
Qy	1874	gcct 1877	
Db	541	gcct 544	
RESULT 5			
LOCUS	AZ661404	545 bp	DNA linear GSS 14-DEC-2000
DEFINITION	IM0540B04F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC1M0540B04 F, DNA sequence.		
ACCESSION	AZ661404		
VERSION	A2661404.1 GI:11798550		
KEYWORDS	GSS.		
SOURCE	house musculus		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 545) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mamoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A., and Wright,D., Weiss,R.		
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT, 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0540 row: B column: 04 Seq primer: CGTTGTAAACGACGCGCAGT Class: plasmid ends High quality sequence stop: 545.		
FEATURES	Location/Qualifiers		
source	1..545 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="UUGC1M0540B04" /clone_lib="Mouse 10kb plasmid UUGC1M library" /sex="Male" /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-" /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWB42 (ql14732114 gb AF129072.1), a copy-number		

[illegible]

Oy 933 ctgtctgggtactgtgactctctccagcgagggaactcttgaacagctgcagctgcaac 992
 Db 361 CTGTCTGGTACTGTGACTGCTTCTCCAGCGGAGACTTCTCAACAGCTGCAGCTGCAAC 420
 Oy 993 aactgcgcctagctgcagctgcgttcaaaagccataaa-gcgtgtcttgcataaaatccc 1051
 Db 421 AACCTGGCCCATGAGCTGAGCGCTTCAAGCCATAAAGGCGTGTCTTGATAGAAATCC 480
 Oy 1052 tgaagctttccaacaaaaatgggg 1076
 Db 481 TGAAGCTTTCCAACAAAATGGGG 505

RESULT 7
 AK015732
 LOCUS
 DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4930509C02:metallothionein-like 5, testis-specific (tesmin), full insert sequence.
 ACCESSION AK015732
 VERSION AK015732.1 GI:12854117
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:4930509C02.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (sites)
 AUTHORS Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 92279253
 PUBMED 10349636
 REFERENCE 2 (sites)
 AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3 (sites)
 AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaiguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4 (sites)
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5 (bases 1 to 2010)
 AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schraml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,

Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamana,K., I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
 Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
 Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCTCGAGTTAATTAATCCCTCCCTCCCTCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES
 source
 1. .2010
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="MGD:MG1:1907904"
 /db_xref="taxon:10090"
 /clone="4930509C02"
 /sex="male"
 /tissue_type="testis"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 67. .900
 /gene="Mt15"
 67. .900
 /gene="Mt15"
 /note="data source:MGD, source key:MG1:1340029, evidence:ISS
 metallothionein-like 5, testis-specific (tesmin)
 putative"
 /codon_start=1
 /protein_id="BAB29949.1"
 /db_xref="GI:12854178"
 /translation="MEDALLGAMTGPEDLFGSERVFDGLALSPAGGAADRD
 LPVLADAYLGATEPEPLRALSPPEALVPAALGDFPGLPEURSPDDRAAPPAYSV
 HVLSLLPGARGPALPLSAGVRVIVPEIKAGGVPGGSPEDAFAOPLAQESCCRF
 PSSQEAEEASSCKDPSPMVICQLGGAQMLCIDNGARELKLHLLPOYDDQSSF
 POSELKPMPTTLVGRLLPVPKLNLTQVDNGALPSAVNGAFAFSGPALQGPRTILS
 G"

polyA_signal
 1991. .1996
 /note="putative"
 2010
 polyA_site
 /note="putative"
 BASE COUNT 407 a 558 c 563 g 482 t
 ORIGIN

Query Match 20.0%; Score 449; DB 11; Length 2010;
 Best Local Similarity 100.0%; Pred. No. 2.le-64;
 Matches 449; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 495 gttgaatacaagaagcagggtgtagtgcaggcgccctgaagacagcagctttc 554
 Db 451 GTTGAATCAAGAAGCAGGTGGTAGTGCCAGCGCCCTGAAGACGACGCTTTC 510
 Oy 555 caggccctctggctcaggaatcctctgtgcaagttcccatcatcccccagagcagagg 614

```

|||||
Db 511 CAGGCCCTCTGGCTCAGGAATCCTGTTCAAGTTCCCATCATCCAGAGGAGGAG 570
QY 615 gctccagctgcccctggaagaagactccagcccatggtgattgtcagctgaaagga 674
|||||
Db 571 GCTCCAGCTGCCCTCGGAAGAAGACTCCAGCCCATGTTGTTGTCAGTGAAGAAG 630
QY 675 ggcgcagatgctctgcatagacaactgtgcgagagagctcaaaacgctccatctg 734
|||||
Db 631 GCGGCCAGATGCTCTGCATAGACAACCTGTGGCGGAGGAGCTCAAGCGCTCCATCTG 690
QY 735 ctctccagtaoatgacagagagcagtttccctcagtcagagagctccctaaagcaatgaca 794
|||||
Db 691 CTTCCTCAGTACCATGACCAAGACAGTCTCCCTCAGTCAGAGCTCCCTAAGCCAATGACA 750
QY 795 actttagtggagactcttcgcagaccaggaaggttaaatctcatcacaggttgat 854
|||||
Db 751 ACTTTAGTGGGAAGACTCTGCCAGTACCAAGCAAGTAAATCTCATCACACAGGTTGAT 810
QY 855 aatggagctctcccatcagctgcaatggggtgccttccctcctggacctgctgcaa 914
|||||
Db 811 AATGGAGCTCTCCCATCAGCTGTCAATGGGCTGCCTTTCCCTCTGGACCTGCTCTGCAA 870
QY 915 gggccaccacaaataactctgtctgggta 943
Db 871 GGGCCACCAAAATAACTCTGTCTGGGTA 899

```

```

RESULT 8
AW558944/c
LOCUS
DEFINITION
L0302B03-3 NIA Mouse Newborn Ovary cDNA Library Mus musculus cDNA
clone L0302B03 3', mRNA sequence.
ACCESSION
AW558944
VERSION
AW558944.1 GI:7204373
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 468)
Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X., Grahovac
M.J., Pantano,S., Sano,Y., Piao,Y., Nagaraia,R., Doi,H., Wood,W.H.
III, Becker,K.G. and Ko,M.S.H.
Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
20381348
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: L0302 row: B column: 03
Seq primer: -21M13 Forward
High quality sequence stop: 468
POLYA=Yes.
Location/Qualifiers
1. 468
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="niaEST:L0302B03-3"
/db_xref="taxon:10090"
/clone="L0302B03"
/clone_lib="NIA Mouse Newborn Ovary cDNA Library"
/sex="female"
/dev_stage="Newborn Ovary"
/lab_host="DH10B"
/notes="vector: pSPORT1 (Gibco/BRL Life Technology);
Site.1: Salt; Site.2: Noti; Total RNAs were extracted from
7 Newborn Ovary. The double-stranded cDNA was synthesized
by Gibco's kit with an Oligo(dT) primer [Noti]
primer-adaptor from GibcoBRL)

```

FEATURES

RESULT 9

```

BF452508
LOCUS
DEFINITION
uz97g12.y1 NCI_CGAP_Tel1 Mus musculus cDNA clone IMAGE:3809495 5',
mRNA sequence.
ACCESSION
BF452508
VERSION
BF452508.1 GI:11518677
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 418)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

```

[5'-pCACTAGTCTAGATCGGAGCGGCCCTTTT-3'] from 2.56ug of total RNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lona-linker LL-Sal3 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Xiaohong Wang and Yulan Piao."

BASE COUNT 146 a 117 c 112 g 93 t
ORIGIN

Query Match 19.6%; Score 440.2; DB 9; Length 468;

Best Local Similarity 98.9%; Pred. No. 1.2e-62;

Matches 464; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1745 ctctctgtccctggggagggccctctgggagctccctgcctgcataaaagaggtgatt 1804

Db 468 CCTCTGTCTCTGGGAGGCGCTCTGGGACTCCCTGCATATAAAGAGGGTGATT 409

QY 1805 ttctactgtgtatgtgttgccttcaaatgcttagtagtacctccattcaagtat 1864

Db 408 TTCTACTTGTGTGTATGTGTGTCTTCAATGTCTTAGTAGTACCTCCATTCGAGTTAT 349

QY 1865 tatgagccagcctcaagttagagagctgcttcttcaggctggactctgccaaatca 1924

Db 348 TATGAGCCAGCCTCAAGTTAGAGAGCTAGGCTCTTCTCAGGTGGACTCTGCCCAATCA 289

QY 1925 catacaagtcaggtggccatcaggggttttccagggccagggcctgacagagatatgg 1984

Db 288 CATACAAAGTCAGGTGCCCATCAGGGG-TCTCCAGGCCAGGCGCTGTGACAGGAGATATGG 230

QY 1985 gaggggggtcggttagagctgggtttgttgatttttgcgttttttcttcctgtat 2044

Db 229 GAGGGGGTCTGGGTTAGAGCTGGGTTTGTGGATTTTTCGCTTTTCTTCCTGTAT 170

QY 2045 ttctgttgaagtggagaaacttgcctcctgtccaa-ctttctccataataattactgctg 2103

Db 169 TTCTGCTTGAAGTGAGAAAACTTGTCTCCTGTCCAAACCTCTTCTTCATTAATTACTGCTG 110

QY 2104 caggtgcctgctgacagtcacagtcacacacacagagtgaggtgcttatt 2163

Db 109 CAGGGTGGCTGTGTGACCATCAGTACAGTACCTCAGACACCCAGAGGTGAGGTGGCTTATT 50

QY 2164 atgccacactttgtgtttgtgtgagaaataaacctttccagactccc 2212

Db 49 ATGCCACACTTTGTGTTGTGTGTGAGAATAAACCTTTCCAGACTCCC 1

JOURNAL
COMMENT

Unpublished (1997)
Other_ESTS: u297g12.x1
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

MGI:1451607

Seq primer: -40RP from Gibco
High quality sequence stop: 381.

FEATURES

source

1. .418
Location/Qualifiers
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3809495"
/clone_lib="NCI_CGAP_Tel"
/sex="male"
/dev_stage="5 months"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.3 kb. Library constructed by Life
Technologies."

BASE COUNT 89 a 101 c 97 g 131 t
ORIGIN

Query Match 16.9%; Score 379.8; DB 10; Length 418;

Best Local Similarity 98.8%; Pred. No. 9.7e-53; Indels 3; Gaps 3;
Matches 414; Conservative 0; Mismatches 2;

Qy 1 tatectgtggtgccccggcagcaggtctcagcaggtcagcagcaccacacagggata 60
Db 2 TATCCTGTGGTGGCCCGGCGCAGGCG-CTCAGCAGGCTCAGGCACACAA-GGATA 59
Qy 61 caccagtgtgttctctggcctgtgtgactgtgtgactccaccacccctccgccccagcagg 120
Db 60 CACAGTGTGTCTTGGCCTGTGGACTTGTGCTCTCCACCACCTCCGCGCCCGCAGGG 119
Qy 121 ctagggtataaacccagggccttttgcgtgtctgcagatagttctcagcctgtagttt 180
Db 120 CTAGGGATAGAACCCAGGGCCCTTTGGCGTGTCTGCAGATAGTTTTCAGCCCTGTAGTTT 179
Qy 181 ggggtgtgctgggaga-tttttttcttcacacaaagacttccattattgaggatttt 239
Db 180 GGGCTTGGCTGGGAGATTTTTTTTCTTACACCAAGACTTCCATATTGAGGATTTT 239
Qy 240 ttcagttgatgatctccccctctgttaagataaggagcagttctttaaacctatgtagag 299
Db 240 TTCAGTTGATGATCTCCCCCTCTGTAAAGATAAGGACAGTCTTTAAACCTATGTAGAG 299
Qy 300 ttttgatgaattctgcttccaaacctatgtcgaagctatataagcaattccttgaattg 359
Db 300 TTTTGTGAAATTCGTCTCAACCATATTGCTAAGCTATATATACAAATTCCTGAAATTG 359
Qy 360 ctataacttaggagaaacctctgattctctcctcctacactcctgagtgtagtgta 418
Db 360 CTATATAACTTAGGAGAACCTCTGATCTCTGCTCTCTACATCTGAGTGTAGTGTA 418

-RESULT 10

B1829802

LOCUS

DEFINITION 603079876F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5171394 5',
842 bp mRNA linear EST 04-OCT-2001

B1829802

mRNA sequence.

ACCESSION

B1829802

VERSION

B1829802.1

KEYWORDS

EST.

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 842)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM11426 row: h column: 19

High quality sequence stop: 821.

Location/Qualifiers

1. .842

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5171394"

/clone_lib="NIH_MGC_119"

/tissue_type="medulla"

/lab_host="DH10B"

/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 013. Note:
this is a NIH_MGC Library."

BASE COUNT 239 a 183 c 207 g 213 t

ORIGIN

Query Match

15.3%; Score 341.8; DB 10; Length 842;

Best Local Similarity 78.2%; Pred. No. 1.2e-46;

Matches 428; Conservative 0; Mismatches 107; Indels 12; Gaps 1;

Qy 946 gtgactgtcttcacagcggggaactctgcacacagctgca-----gctgcaaca 993
Db 1 GTGACTGCTTTGCCAGTGGGGACGTTTGCACAACTTGCAATTTGTAATTAATTTGTCACAA 60
Qy 994 acctgcccagctgagctgcagcgtctcaaacataaacgagcgtgctctgtagaatactcg 1053
Db 61 ACTTGCAATCATGATATTGAACGGTTTAAAGCCATTAAAGCCATTAGGCATGCTTGGTAGAATCCAG 120
Qy 1054 aagctttccaaacaaaaatgggaaagccctctgggagctgctaaacttcgacacagca 1113
Db 121 AAGCTTCCAGCCAAAATTTGGGAAGGGCCCAATTTGGGCAATGTCAGCCCGCCAGCAACA 180
Qy 1114 aagggtgcaactgaagcgtcagcgtcagcgtcctgaagaactactgtgagtgatagggca 1173
Db 181 AAGGTGCAACTGAGGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 240
Qy 1174 aaatcatgtgtcttcctcattgcaaatgcatgctgtcgaactatgaagaagtcag 1233
Db 241 AAATTATGTGTTCTTCTATTGCAATTCATTTGGTGGCAAAATTTATGCAAAAGCCAG 300
Qy 1234 aacgaaaaatgtgatgagcacacccccacacacacacacacacacacacacacacacacacac 1293
Db 301 AACGAAAGACACTAATGAGCATGCAAACTACATGCAGACTGGAGGTTTGGAGGACGCC 360
Qy 1294 attattgtccccagcgaagttctcaggacacctcaaaactgagaaaaaataggagcgt 1353
Db 361 ATTACCTGCCACCAACGAATTTTTCAGGACTTCCAAAGATTCCAGATTCAGATAGGGCGCTT 420
Qy 1354 tctcctgtatctcctgggaagtagtgaggagccacatgtgctgctgctgctgctgctgctgctg 1413

Db 192 AGTGACCTCAGACACAGAAAGGTGAGGTGGCTATTATGCCACACTTGTGTTTGTG 251

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 08:14:43 ; Search time 95.08 seconds
(without alignments)
5789.490 Million cell updates/sec

Title: US-09-743-237-1
Perfect score: 2241
Sequence: 1 tatcctgtggttgcccg.....aaaaaaaaaaaaaaaaaaaa 2241

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	44.2	2.0	615	4	US-09-328-111-371	Sequence 371, App
C 2	43.6	1.9	964	1	US-08-448-606-5	Sequence 5, Appl
C 3	43	1.9	658	4	US-09-385-982-327	Sequence 327, App
C 4	42.8	1.9	3024	6	5284931-1	Patent No. 5284931
C 5	41	1.8	1129	4	US-09-227-357-40	Sequence 40, Appl
C 6	40.4	1.8	333	3	US-08-946-026-35	Sequence 35, Appl
C 7	40.4	1.8	578	4	US-09-328-111-757	Sequence 757, App
C 8	40.4	1.8	1977	4	US-09-227-357-83	Sequence 83, Appl
C 9	40.4	1.8	2238	1	US-08-742-011-1	Sequence 1, Appl
C 10	40.4	1.8	7218	1	US-08-232-463-14	Sequence 14, Appl
C 11	40.2	1.8	672	1	US-07-847-010-21	Sequence 21, Appl
C 12	40.2	1.8	1181	1	US-08-181-271A-27	Sequence 27, Appl
C 13	40.2	1.8	1181	1	US-08-449-315-27	Sequence 27, Appl
C 14	40.2	1.8	1181	1	US-08-444-803-27	Sequence 27, Appl
C 15	40.2	1.8	1181	1	US-08-449-043-27	Sequence 27, Appl
C 16	40.2	1.8	1181	1	US-08-456-265A-27	Sequence 27, Appl
C 17	40.2	1.8	1181	1	US-08-456-416-27	Sequence 27, Appl
C 18	40.2	1.8	1181	1	US-08-455-244-27	Sequence 27, Appl
C 19	40.2	1.8	1181	1	US-08-454-876-27	Sequence 27, Appl
C 20	40.2	1.8	1181	2	US-08-457-364-27	Sequence 27, Appl
C 21	40.2	1.8	1181	2	US-08-456-262-27	Sequence 27, Appl
C 22	40.2	1.8	1181	2	US-08-456-240-27	Sequence 27, Appl
C 23	40.2	1.8	1181	2	US-08-455-736-27	Sequence 27, Appl
C 24	40.2	1.8	1181	2	US-08-971-217-27	Sequence 27, Appl
C 25	40.2	1.8	1181	4	US-09-350-600-27	Sequence 27, Appl
C 26	40.2	1.8	2672	1	US-08-703-947-1	Sequence 1, Appl
C 27	39.4	1.8	903	5	PCT-US95-06406A-21	Sequence 21, Appl

28	38.6	1.7	882	2	US-08-909-965C-9	Sequence 9, Appl
29	38.6	1.7	1733	3	US-09-073-569-1	Sequence 1, Appl
30	38.6	1.7	3217	4	US-09-232-200-64	Sequence 64, Appl
31	38.6	1.7	3217	4	US-09-232-137-64	Sequence 64, Appl
32	38.6	1.7	3217	4	US-09-232-201-64	Sequence 64, Appl
33	38.6	1.7	3471	5	PCT-US93-00227-2	Sequence 2, Appl
34	38.4	1.7	770	3	US-08-938-675A-1	Sequence 1, Appl
35	38.4	1.7	1858	2	US-08-909-965C-11	Sequence 11, Appl
36	38.4	1.7	2409	4	US-09-293-322C-8	Sequence 8, Appl
37	38.4	1.7	152331	3	US-09-128-155-16	Sequence 16, Appl
C 38	38.2	1.7	658	4	US-09-328-111-816	Sequence 816, App
39	38.2	1.7	770	3	US-08-865-297-5	Sequence 5, Appl
40	38.2	1.7	2107	4	US-09-180-852-1	Sequence 1, Appl
41	38.2	1.7	2413	4	US-09-518-046-1	Sequence 1, Appl
42	38.2	1.7	2416	4	US-09-261-416-1	Sequence 1, Appl
43	38.2	1.7	2544	4	US-09-518-046-3	Sequence 3, Appl
44	38.2	1.7	3761	4	US-08-890-865A-2	Sequence 2, Appl
45	38	1.7	374	2	US-08-370-156-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-09-328-111-371/c
; Sequence 371, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 371
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(615)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-371

Query Match 2.0%; Score 44.2; DB 4; Length 615;
Best Local Similarity 75.3%; Pred. No. 0.0055;
Matches 55; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 2169 cacacttggttgttgtagaataaacctttccagactcccaaaaaaaaaaaaaa 2228

Db 82 CAGACTTTGTGTATTTGTTATTAAGCATTCCTTAACAGAAAAAAAAAAAAA 23

Qy 2229 aaaaaaaaaaaaaa 2241

Db 22 AAAAAAAAAAAAAA 10

RESULT 3
US-09-385-982-327/C

Query Match 1.8%; Score 40.4; DB 3; Length 333;
Best Local Similarity 68.3%; Pred. No. 0.047;
Matches 56: Conservative 0; Mismatches 26: Indels

/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/051,929	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/052,803	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/052,732	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/051,931	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/051,932	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/051,916	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/051,930	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/051,918	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/051,920	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/052,733	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/052,795	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/051,919	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/051,928	
/	EARLIER FILING DATE:	1997-07-08	
/	EARLIER APPLICATION NUMBER:	60/055,722	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/055,723	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/055,948	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/055,949	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/055,953	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/055,950	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/055,947	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/055,964	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/056,360	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/055,684	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/055,984	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/055,954	
/	EARLIER FILING DATE:	1997-08-18	
/	EARLIER APPLICATION NUMBER:	60/058,785	
/	EARLIER FILING DATE:	1997-09-12	
/	EARLIER APPLICATION NUMBER:	60/058,664	
/	EARLIER FILING DATE:	1997-09-12	
/	EARLIER APPLICATION NUMBER:	60/058,660	
/	EARLIER FILING DATE:	1997-09-12	
/	EARLIER APPLICATION NUMBER:	60/058,661	
/	EARLIER FILING DATE:	1997-09-12	
/	NUMBER OF SEQ ID NOS:	672	
/	SOFTWARE:	PatentIn Ver. 2.0	
/	SEQ ID NO 83		
/	LENGTH:	1977	
/	TYPE:	DNA	
/	ORGANISM:	Homo sapiens	
/	FEATURE:		
/	NAME/KEY:	SITE	
/	LOCATION:	{664}	
/	OTHER INFORMATION:	n equals a,t,g, or c	
/	FEATURE:		
/	NAME/KEY:	SITE	
/	LOCATION:	{716}	
/	OTHER INFORMATION:	n equals a,t,g, or c	

RESULT 7
US-09-328-111-757/c

```

/ EARLIER FILING DATE: 1997-07-08
/ EARLIER APPLICATION NUMBER: 60/051,930
/ EARLIER FILING DATE: 1997-07-08
/ EARLIER APPLICATION NUMBER: 60/051,918
/ EARLIER FILING DATE: 1997-07-08
/ EARLIER APPLICATION NUMBER: 60/051,920
/ EARLIER FILING DATE: 1997-07-08
/ EARLIER APPLICATION NUMBER: 60/052,733
/ EARLIER FILING DATE: 1997-07-08
/ EARLIER APPLICATION NUMBER: 60/052,795
/ EARLIER FILING DATE: 1997-07-08
/ EARLIER APPLICATION NUMBER: 60/051,919
/ EARLIER FILING DATE: 1997-07-08
/ EARLIER APPLICATION NUMBER: 60/051,928
/ EARLIER FILING DATE: 1997-07-08
/ EARLIER APPLICATION NUMBER: 60/055,722
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/055,723
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/055,948
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/055,949
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/055,953
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/055,950
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/055,947
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/055,964
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/056,360
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/055,684
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/055,984
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/055,954
/ EARLIER FILING DATE: 1997-08-18
/ EARLIER APPLICATION NUMBER: 60/058,785
/ EARLIER FILING DATE: 1997-09-12
/ EARLIER APPLICATION NUMBER: 60/058,664
/ EARLIER FILING DATE: 1997-09-12
/ EARLIER APPLICATION NUMBER: 60/058,660
/ EARLIER FILING DATE: 1997-09-12
/ EARLIER APPLICATION NUMBER: 60/058,661
/ EARLIER FILING DATE: 1997-09-12
/ NUMBER OF SEQ ID NOS: 672
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 83
/ LENGTH: 1977
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (664)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: SITE
/ LOCATION: (716)
/ OTHER INFORMATION: n equals a,t,g, or c

```

```

1 EARLIER FILING DATE: 1997-08-18
2 EARLIER APPLICATION NUMBER: 60/055,984
3 EARLIER FILING DATE: 1997-08-18
4 EARLIER APPLICATION NUMBER: 60/055,954
5 EARLIER FILING DATE: 1997-08-18
6 EARLIER APPLICATION NUMBER: 60/058,785
7 EARLIER FILING DATE: 1997-09-12
8 EARLIER APPLICATION NUMBER: 60/058,664
9 EARLIER FILING DATE: 1997-09-12
10 EARLIER APPLICATION NUMBER: 60/058,660
11 EARLIER FILING DATE: 1997-09-12
12 EARLIER APPLICATION NUMBER: 60/058,661
13 EARLIER FILING DATE: 1997-09-12
14 NUMBER OF SEQ ID NOS: 672
15 SOFTWARE: PatentIn Ver. 2.0
16 SEQ ID NO 83
17 LENGTH: 1977
18 TYPE: DNA
19 ORGANISM: Homo sapiens
20 FEATURE:
21 NAME/KEY: SITE
22 LOCATION: (5664)
23 OTHER INFORMATION: n equals a,t,g, or c
24 FEATURE:
25 NAME/KEY: SITE
26 LOCATION: (7116)
27 OTHER INFORMATION: n equals a,t,g, or c

```

```

1 EARLIER FILING DATE: 1997-09-12
2 EARLIER APPLICATION NUMBER: 60/058,785
3 EARLIER FILING DATE: 1997-09-12
4 EARLIER APPLICATION NUMBER: 60/058,664
5 EARLIER FILING DATE: 1997-09-12
6 EARLIER APPLICATION NUMBER: 60/058,660
7 EARLIER FILING DATE: 1997-09-12
8 EARLIER APPLICATION NUMBER: 60/058,661
9 EARLIER FILING DATE: 1997-09-12
10 NUMBER OF SEQ ID NOS: 672
11 SOFTWARE: PatentIn Ver. 2.0
12 SEQ ID NO 83
13 LENGTH: 1977
14 TYPE: DNA
15 ORGANISM: Homo sapiens
16 FEATURE:
17 NAME/KEY: SITE
18 LOCATION: (664)
19 OTHER INFORMATION: n equals a,t,g, or c
20 FEATURE:
21 NAME/KEY: SITE
22 LOCATION: (716)
23 OTHER INFORMATION: n equals a,t,g, or c

```

RESULT 8
US-09-227-357-83

```

: EARLIER APPLICATION NUMBER: 007058,601
:
: EARLIER FILING DATE: 1979-09-12
:
: NUMBER OF SEQ ID NOS: 672
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 83
:
: LENGTH: 1977
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: FEATURE:
:
: NAME/KEY: SITE
:
: LOCATION: (664)
:
: OTHER INFORMATION: n equals a,t,g, or c
:
: FEATURE:
:
: NAME/KEY: SITE
:
: LOCATION: (716)
:
: OTHER INFORMATION: n equals a,t,g, or c
:

```

FEATURE:
NAME/KEY: SITE
LOCATION: (1319)
OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-83

Query Match 1.8%; Score 40.4; DB 4; Length 1977;
Best Local Similarity 75.8%; Pred. No. 0.13;
Matches 50; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2176 tctgttttggagagataaacctttccagactcccaaaaaaaaaaaaaaaaaa 2235
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1876 tctgtttggagagagataaatgttttcttcaaaaaaaaaaaaaaaaaa 1935

QY 2236 aaaaaa 2241
|||||
Db 1936 aaaaaa 1941

RESULT 9
US-08-742-011-1
; Sequence 1, Application US/08742011
; Patent No. 5824504
; GENERAL INFORMATION:
; APPLICANT: Elshourbagy, Nabil A.
; APPLICANT: Bergsma, Derk J.
; APPLICANT: Ellis, Catherine E.
; TITLE OF INVENTION: Human 7-transmembrane Receptor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742.011
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T.
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50020P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2238 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 332..1357
US-08-742-011-1

Query Match 1.8%; Score 40.4; DB 1; Length 2238;
Best Local Similarity 68.3%; Pred. No. 0.14;
Matches 56; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 2160 tattatgccacacttctgtttttgttgagaataaacctttccagactcccaaaaaa 2219
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2147 TATTACTAGCATATGAGTTTCATAGCTAAGAAATAAACTGTTAAAGTCTCCCAAAAAA 2206

QY 2220 aaaaaaaaaaaaaaaaaa 2241
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2207 AAAAAAAAAAAAAAAAAA 2228

RESULT 10
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 1.8%; Score 40.4; DB 1; Length 7218;
Best Local Similarity 6.1%; Pred. No. 0.28; Mismatches 161; Indels 0; Gaps 0;
Matches 23; Conservative 190

QY 1369 gggaaagttagtgagccacatgtgctgctgctgcccagggtaggaagcagagcagg 1428
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1395 RRR 1336
QY 1429 agcactgttcccaagcttgctgagcagatgatcctggagagagtttgaaagtgctgt 1488
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1335 RRR 1276
QY 1489 cgcagattctcacatcagtgatccaagggcctgaaattgagtagctgcaagct 1548
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1275 RRR 1216
QY 1549 gftaaaggggaatgcctgtggcaagcctcagccctggggaatctgcaccgaggaagctg 1608

Query Match 1.8%; Score 40.2; DB 1; Length 1181;
Best Local Similarity 78.7%; Pred. No. 0.11;
Matches 48; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

2181 ttgtgtgagataaaccctttccagactcccaaaaaaaaaaaaaaaaa 2240
- - - - -
78 TGTATTAATAGAGTAAATCTTCCTTGTCTCAAAAAAAAAAAAAAAAA 19

Qy 2241 a 2241

Db 18 A 18

RESULT 13

RESULT IS
US-08-449-315-27/c
; Sequence 27, Application US/08449315
; Patent No. 5650505

GENERAL INFORMATION:

APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.

1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 08:15:13 ; Search time 5652.01 Seconds
(without alignments)
8578.153 Million cell updates/sec

Title: US-09-743-237-1
Perfect score: 2241
Sequence: 1 tatctctgtggtggcccg.....aaaaaaaaaaaaaaaaaaaaa 2241

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 21979536 seqs, 10817449327 residues

Total number of hits satisfying chosen parameters: 43959072

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*

2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*

5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*

6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*

7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*

8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*

9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*

10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*

11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*

12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*

13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*

14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*

15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*

16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*

17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*

18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*

19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*

24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*

25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*

26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*

27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*

28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*

30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*

32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*

33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*

34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*

35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*

36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*

37: /cgn2_6/ptodata/2/pna/US100_COMB.seq:*

38: /cgn2_6/ptodata/2/pna/US101_COMB.seq:*

39: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*

40: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*

41: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*

42: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*

43: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*

60: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*

61: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*

62: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*

63: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*

64: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*

65: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*

66: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

67: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*

68: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*

69: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*

70: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*

71: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

72: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*

73: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*

74: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*

75: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2241	100.0	2241	29	US-09-743-237-1
2	2066.4	92.2	2201	75	US-60-360-207-983
3	1747	78.0	1861	29	US-09-743-237-2
4	649.4	29.0	2638	25	US-09-643-893-287
5	647.8	28.9	2134	29	US-09-743-237-3
6	560.4	25.0	1770	18	US-09-471-275-5610
7	271.8	12.1	397	16	US-09-289-768-13239
8	271.8	12.1	397	35	US-09-939-397-13239
9	248.4	11.1	445	1	PCT-US01-01329-425
10	248.4	11.1	445	1	PCT-US01-01339-1301
11	248.4	11.1	445	30	US-09-764-891-1301
12	248.4	11.1	445	37	US-10-080-090-425
13	218	9.7	431	13	US-08-971-197-11928
14	218	9.7	431	13	US-08-971-197C-11928
15	175.4	7.8	4276	75	US-60-360-207-19806
16	168	7.5	3495	23	US-09-611-526-2894
17	157.4	7.0	3535	14	US-09-087-136-16
18	157.4	7.0	3535	14	US-09-087-136-16
19	157.4	7.0	3535	16	US-09-220-091-16
20	152	6.8	352	19	US-09-515-128-15843
21	152	6.8	572	16	US-09-205-070-7224
22	152	6.8	572	17	US-09-340-623-7224
23	152	6.8	572	33	US-09-898-888-7224
24	152	6.8	572	33	US-09-898-888A-7224
25	152	6.8	795	1	PCT-US01-08656-5196
26	150.4	6.7	352	18	US-09-471-275-7634
27	123.6	5.5	444	18	US-09-404-549-314
28	123.6	5.5	444	18	US-09-404-549A-314
29	120.4	5.4	250	1	PCT-US00-26524B-2645
30	109.4	4.9	249	20	US-09-539-806-21625
31	105.4	4.7	1093	1	PCT-US01-08631-17949

32 103.8 4.6 465 17 US-09-362-510-45961 Sequence 45961, A
33 103.8 4.6 465 17 US-09-362-510A-45961 Sequence 45961, A
34 103.8 4.6 465 34 US-09-904-013-45961 Sequence 45961, A
35 92.8 4.1 2984 55 US-60-167-217-12947 Sequence 12947, A
36 92.8 4.1 3164 23 US-09-614-150-12896 Sequence 12896, A
37 92.8 4.1 3164 58 US-60-191-637-12934 Sequence 12934, A
38 92.4 4.1 302 14 US-09-013-812-1497 Sequence 1497, Ap
39 92.4 4.1 302 21 US-09-540-764-42635 Sequence 42635, A
40 92.4 4.1 302 42 US-60-036-570-1497 Sequence 1497, Ap
41 78.2 3.5 1716 28 US-09-708-427-33364 Sequence 33364, A
42 78.2 3.5 1812 28 US-09-708-427-27293 Sequence 27293, A
43 76.6 3.4 124 25 US-09-643-893-39 Sequence 39, App1
44 74.8 3.3 97814 20 US-09-534-859-253 Sequence 253, App
45 74.8 3.3 97814 31 US-09-803-736-253 Sequence 253, App

ALIGNMENTS

RESULT 1
US-09-743-237-1
; Sequence 1, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YOUNI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743,237
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2241
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (651)..(1535)
US-09-743-237-1

Query Match 100.0%; Score 2241; DB 29; Length 2241;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 tatectgtggttgccccgggcagcaggctctcagcaggctcaggcaccacacaggata 60
Db 1 tatectgtggttgccccgggcagcaggctctcagcaggctcaggcaccacacaggata 60
Qy 61 cacagtgtgttctctggttggctgtgactccaccaccctccgccccagcagg 120
Db 61 cacagtgtgttctggttggctgtgactccaccaccctccgccccagcagg 120
Qy 121 ctaggatagaacccaggccttttgcgtgtctgcagatgtctcagcgtgagttt 180
Db 121 ctaggatagaacccaggccttttgcgtgtctgcagatgtctcagcgtgagttt 180
Qy 181 ggggttggtggagattttttttttttttcacacaaagacttccattattgagatttt 240
Db 181 ggggttggtggagattttttttttttttcacacaaagacttccattattgagatttt 240
Qy 241 tcagttgatgatccccctctgttaagataaaggacagttcttttaaacctatgagagt 300
Db 241 tcagttgatgatccccctctgttaagataaaggacagttcttttaaacctatgagagt 300
Qy 301 ttgatgaattctgtctcaaccattattgactaatatagcaattcttgaattgc 360
Db 301 ttgatgaattctgtctcaaccattattgactaatatagcaattcttgaattgc 360

Qy 361 tatataaacttagagagaacctctgattctctctgctctacatcctgactgactaggtgtaca 420
Db 361 tatataaacttagagagaacctctgattctctctgctctacatcctgactgactaggtgtaca 420
Qy 421 gggggaatacatttgggtgagactccgatgaactactgcaggttcccaagcagcaagc 480
Db 421 gggggaatacatttgggtgagactccgatgaactactgcaggttcccaagcagcaagc 480
Qy 481 aagcaagaaaaagtgttgaatacaagaagcagggtggttagtggtgcagcgccgacctg 540
Db 481 aagcaagaaaaagtgttgaatacaagaagcagggtggttagtggtgcagcgccgacctg 540
Qy 541 aagacgcagctttccaggccctctggtctcaggaatctctgttgcaggttcccatatccc 600
Db 541 aagacgcagctttccaggccctctggtctcaggaatctctgttgcaggttcccatatccc 600
Qy 601 aggaggcagagaggcctccagctgccctcgaaagaagactccagcccatgggtgattt 660
Db 601 aggaggcagagaggcctccagctgccctcgaaagaagactccagcccatgggtgattt 660
Qy 661 gtcagctgaagaggcgcgccagatgctctgcatagaacaactgtggtgcgagggagctca 720
Db 661 gtcagctgaagaggcgcgccagatgctctgcatagaacaactgtggtgcgagggagctca 720
Qy 721 aagcgtccatctgttctcagtaacgatgaccagagcaggtttccctcagtcagagctcc 780
Db 721 aagcgtccatctgttctcagtaacgatgaccagagcaggtttccctcagtcagagctcc 780
Qy 781 ctaagccaatgacaacttttagtgggaagactctgcagtcaccagcgaagttaaaactca 840
Db 781 ctaagccaatgacaacttttagtgggaagactctgcagtcaccagcgaagttaaaactca 840
Qy 841 tcacacaggttgataatggagctctcccatcagctgcaatggggtgcttccctctcg 900
Db 841 tcacacaggttgataatggagctctcccatcagctgcaatggggtgcttccctctcg 900
Qy 901 gacctgtctgcaaggccaccacaaataaactctgtctgggtactgtgactgctcttcca 960
Db 901 gacctgtctgcaaggccaccacaaataaactctgtctgggtactgtgactgctcttcca 960
Qy 961 gggggactctgcacaagctgcagctgcaacaacctgcgcatgagctgcagcgtctca 1020
Db 961 gggggactctgcacaagctgcagctgcaacaacctgcgcatgagctgcagcgtctca 1020
Qy 1021 aagccataaaggcgtgtctgtatagaaatctctgaaagctttccaaccaaataagggaag 1080
Db 1021 aagccataaaggcgtgtctgtatagaaatctctgaaagctttccaaccaaataagggaag 1080
Qy 1081 ggcgtctggagcgtgctaaacttcgacacagcaaggggtgcaactgaagcgtcaggct 1140
Db 1081 ggcgtctggagcgtgctaaacttcgacacagcaaggggtgcaactgaagcgtcaggct 1140
Qy 1141 gctggaagactactgtgagtgctatgaggcgaataatcatgtgttctccattgcaat 1200
Db 1141 gctggaagactactgtgagtgctatgaggcgaataatcatgtgttctccattgcaat 1200
Qy 1201 gcatgtctgcaaaactatgaagaagtccagaaagaaatgctgtagagcacacccc 1260
Db 1201 gcatgtctgcaaaactatgaagaagtccagaaagaaatgctgtagagcacacccc 1260
Qy 1261 actaatgagcctgggacttgagagcagccattattgtccccagcccaattctcag 1320
Db 1261 actaatgagcctgggacttgagagcagccattattgtccccagcccaattctcag 1320
Qy 1321 gacctccaaactgagaaaaatagcagcgccttctctgtatctctctgggaagtagtg 1380
Db 1321 gacctccaaactgagaaaaatagcagcgccttctctgtatctctctgggaagtagtg 1380
Qy 1381 aggccacatgtgctgctgctgctggccagggtgaggaagcagagcaggtgactgttccc 1440
Db 1381 aggccacatgtgctgctgctgctggccagggtgaggaagcagagcaggtgactgttccc 1440
Qy 1441 caagcttgctgagcagatgatcctggaggaggttgggaaggtgcctgtgcgagattctcc 1500

```

1441 caagcttgctgagcagatgatctctgagggagttgtggaagtgctgctgcagattctcc 1500
1501 acatcgaattcaagctcaagggctgaaattgtagtagctgcaagctggttaaaggggaa 1560
1501 acatcgagttcaagctcaagggctgaaattgtagtagctgcaagctggttaaaggggaa 1560
1561 tgcctgtggcaagctcagcctcggaatctgcaccaggaagctggtgcaccagggagga 1620
1561 tgcctgtggcaagctcagcctcggaatctgcaccaggaagctggtgcaccagggagga 1620
1621 gcaaggccgcgcacatcagccaggtcagctgtgaggtctgtagtctgtagtgcattgtagt 1680
1621 gcaaggccgcgcacatcagccaggtcagctgtgaggtctgtagtctgtagtgcattgtagt 1680
1681 gccagctactcaaggtatctcaagtcgaagcagcagcagcagcagcagcagcagcagcagc 1740
1681 gccagctactcaaggtatctcaagtcgaagcagcagcagcagcagcagcagcagcagcagc 1740
1741 tggccctctgctcctgggagccctctgggagctcctcgccctgcataaaagaggggt 1800
1741 tggccctctgctcctgggagccctctgggagctcctcgccctgcataaaagaggggt 1800
1801 gattttctactgtgttatgtgttgccttcaaatgtcttagtagtacctccattcaag 1860
1801 gattttctactgtgttatgtgttgccttcaaatgtcttagtagtacctccattcaag 1860
1861 ttattatgagccagcccaatgagagctagcgtctcttctcaggtgagctgcgccaa 1920
1861 ttattatgagccagcccaatgagagctagcgtctcttctcaggtgagctgcgccaa 1920
1921 atcacatacaagtcaggtggccatcaggggttttccagcagccctgtgacagggagat 1980
1921 atcacatacaagtcaggtggccatcaggggttttccagcagccctgtgacagggagat 1980
1981 atgggaggggggttcgggttagagctgggtttgttggatttttgcgtttttttcttct 2040
1981 atgggaggggggttcgggttagagctgggtttgttggatttttgcgtttttttcttct 2040
2041 gtatttctgctgaagtagagaaactgtctcctgtccaaacttttctccataattactg 2100
2041 gtatttctgctgaagtagagaaactgtctcctgtccaaacttttctccataattactg 2100
2101 ctgcacgtgcctgctgaccagtcacagtcacctcagacaccagaaagtgagtggtt 2160
2101 ctgcacgtgcctgctgaccagtcacagtcacctcagacaccagaaagtgagtggtt 2160
2161 attatgccacacttctgtgtttgttggtagaataaaactttccagactcccaaaaaa 2220
2161 attatgccacacttctgtgtttgttggtagaataaaactttccagactcccaaaaaa 2220
2221 aaaaaaaaaaaaaaaaaa 2241
2221 aaaaaaaaaaaaaaaaaa 2241
```

RESULT 2

```

US-60-360-207-983
; Sequence 983, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60360, 207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 983
; LENGTH: 2201
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-983
```

```

Query Match          92.2%; Score 2066.4; DB 75; Length 2201;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 2170; Conservative 0; Mismatches 16; Indels 14; Gaps 8;

Qy 20 gggcagcaggctctcagcaggctcagccaccacacagggatcacacagtggtgttctctggc 79
Db 9 gggctggcccgagcaggctcagccaccacacagggatcacacagtggtgttctctggc 67
Qy 80 ctgttgacttgtagctccaccaccctcccccagcagggctagggatagaaacccagg 139
Db 68 ctgttgacttgtagctccaccaccctcccccagcagggctagggatagaaacccagg 127
Qy 140 cctttgctgttctcagatagctctcagcctgtgagtttgggttggctgaggagatt 199
Db 128 cctttgctgttctcagatagctctcagcctgtgagtttgggttggctgaggagatt 187
Qy 200 tttttctcacacaaagacttccattattgaggtatttttcagttgtagatctcccc 259
Db 188 tttttctcacacaaagacttccattattgaggtatttttcagttgtagatctcccc 246
Qy 260 ctctgtaagataagggacagttctttaaaccctatgtagagttttgagaaattctgcttct 319
Db 247 ctctgtaaga caaggagcagttctttaaaccctatgtagagtttttcagttgtagatctct 306
Qy 320 caacatattgctaaagctatatagcaattctctgaaattgctataataacttaggagaacc 379
Db 307 caacatattgctaaagctatatagcaattctctgaaattgctataataacttaggagaacc 366
Qy 380 tctgattctcctcctctacatcctgagtgctaggtgtacaggggggaaatcttttgggtg 439
Db 367 tctgattctcctcctctacatcctgagtgctaggtgtacaggggggaaatcttttgggtg 426
Qy 440 agactcagatgaactactgcccagttcccaggcagcaagcaagcaagcaagcaagcaagtttga 499
Db 427 agactcgaatgaactactgcccagttcccaggcagcaagcaagcaagcaagcaagtttga 482
Qy 500 aatcaagaagcaggtggttagtggtgagcagcctgaaagcagcagctttccagggc 559
Db 483 aatcaagaagcaggtggttagtggtgagcagcctgaaagcagcagctttccagggc 542
Qy 560 cctctggtcaggaatctctgttcaagttcccatcatcccagagggcagaggggacctc 619
Db 543 cctctggtcaggaatctctgttcaagttcccatcatcccagagggcagaggggacctc 602
Qy 620 cagctcctcaggaagaaagactccagcccatggtgattgttcagctgaaagagggcgc 679
Db 603 cagctcctcaggaagaaagactccagcccatggtgattgttcagctgaaagagggcgc 662
Qy 680 ccagatgctctgcataagaaactgtgagcagggagcagcagcagcagcagcagcagcagcagc 739
Db 663 ccagatgctctgcataagaaactgtgagcagggagcagcagcagcagcagcagcagcagcagc 722
Qy 740 tcagtacgatgacacagcagcttccctcagtcagagctcctcagcagcagcagcagcagcagcagc 799
Db 723 tcagtacgatgacacagcagcttccctcagtcagagctcctcagcagcagcagcagcagcagcagc 782
Qy 800 agtgggaagacttctgcagtcaccaggaagttaaatctcaccacacaggttgataaagg 859
Db 783 agtgggaagacttctgcagtcaccaggaagttaaatctcaccacacaggttgataaagg 842
Qy 860 agctctcccatcagctgcaatggggctgctttccctcctgagcctgctcctgcaggggccc 919
Db 843 agctctcccatcagctgcaatggggctgctttccctcctgagcctgctcctgcaggggccc 902
Qy 920 acccaaaataactctgtctgggtactgtgactgtcttccagcgggggacttctgcacag 979
Db 903 acccaaaataactctgtctgggtactgtgactgtcttccagcgggggacttctgcacag 962
Qy 980 ctgcagctgcaacacactgcacctgagctgagcagcagcagcagcagcagcagcagcagcagc 1039
Db 963 ctgcagctgcaacacactgcacctgagctgagcagcagcagcagcagcagcagcagcagcagc 1022
Qy 1040 tgatagaaatcctgaagctttcccaacaaatgggggaagggcgctctgggagcgtctaa 1099
```


QY 1035 tgccttgatagaatcctggaactttccaaacccaaatgggaaagccgtctggagct 1094
 Db 655 tgccttgatagaatcctggaactttccaaacccaaatgggaaagccgtctggagct 714
 QY 1095 gctaaacttcgacagcagaaaggggtgaactgtaagcgtcaggtgctgaagaactac 1154
 Db 715 gctaaacttcgacagcagaaaggggtgaactgtaagcgtcaggtgctgaagaactac 774
 QY 1155 tgcgtgctatgagggccaaatcatgtgtcttccattgcaaatgcatgtctgcaaa 1214
 Db 775 tgcgtgctatgagggccaaatcatgtgtcttccattgcaaatgcatgtctgcaaa 834
 QY 1215 aactatgaagaagtcacgaacgaataatgctgatgagcacaccaccactacatgagcct 1274
 Db 835 aactatgaagaagtcacgaacgaataatgctgatgagcacaccaccactacatgagcct 894
 QY 1275 ggggactttgagagcagcattattgttcccccagcagaagttctcaggacctccaaactg 1334
 Db 895 ggggactttgagagcagcattattgttcccccagcagaagttctcaggacctccaaactg 954
 QY 1335 agaaaaatagcagcagccttccctgtatctctctggaagtagtgagggccacatgtgcc 1394
 Db 955 agaaaaatagcagcagccttccctgtatctctctggaagtagtgagggccacatgtgcc 1014
 QY 1395 tgcctgctggccaggggtgagaaagcagagcagggagcagctgtcccccagcttggctgag 1454
 Db 1015 tgcctgctggccaggggtgagaaagcagagcagggagcagctgtcccccagcttggctgag 1074
 QY 1455 cagatgactcctggagaggttggaggggtgctgtgcagagattctccacatcgagttcaag 1514
 Db 1075 cagatgactcctggagaggttggaggggtgctgtgcagagattctccacatcgagttcaag 1134
 QY 1515 tccaggggctgaataatgagtagcgtgcaagctgtaaggggaatgccttgcaagc 1574
 Db 1135 tccaggggctgaataatgagtagcgtgcaagctgtaaggggaatgccttgcaagc 1194
 QY 1575 ctccagccttggaaatctgcacagaggaagctgtgtccagggagagcagagggccgcga 1634
 Db 1195 ctccagccttggaaatctgcacagaggaagctgtgtccagggagagcagagggccgcga 1254
 QY 1635 tcatggcaggtcagctgtgaggtctgagtgatctgcatgtgtactggccagcactactcaa 1694
 Db 1255 tcatggcaggtcagctgtgaggtctgagtgatctgcatgtgtactggccagcactactcaa 1314
 QY 1695 ggtatcttaagtgcaagcagcagacacccctgggaatggaactggcctctctgtcc 1754
 Db 1315 ggtatcttaagtgcaagcagcagacacccctgggaatggaactggcctctctgtcc 1374
 QY 1755 ctggggaggccctctgggactccctgcccctgcataaaaaagggtgattttctacttgt 1814
 Db 1375 ctggggaggccctctgggactccctgcccctgcataaaaaagggtgattttctacttgt 1434
 QY 1815 tgttatgtttgtcttcaaatgtttagtagtagtaactccatcaagtattattagccag 1874
 Db 1435 tgttatgtttgtcttcaaatgtttagtagtagtaactccatcaagtattattagccag 1494
 QY 1875 cctcaagttaagagtaggctctctcagtgagtgagctgcccacaaatcaacataagtc 1934
 Db 1495 cctcaagttaagagtaggctctctcagtgagtgagctgcccacaaatcaacataagtc 1554
 QY 1935 aggtggccatcaggggttttccagggcagggcctgtgacagagatagggggggggtc 1994
 Db 1555 aggtggccatcaggggttttccagggcagggcctgtgacagagatagggggggggtc 1614
 QY 1995 ggggttagagctgggtttgttggatttttgggttttttttttctctgtatttctgtga 2054
 Db 1615 ggggttagagctgggtttgttggatttttgggttttttttttctctgtatttctgtga 1674
 QY 2055 agtgagaaactgtctctgtccaaactttctcctaattactgtgcagggctgcct 2114
 Db 1675 agtgagaaactgtctctgtccaaactttctcctaattactgtgcagggctgcct 1734
 QY 2115 gctgaccagtcagctgacccctcagacacagagggtgaggtggttattatgcccacact 2174

Db 1735 gctgaccagtcagctgacccctcagacacagagggtgaggttattatgcccacact 1794
 QY 2175 ttcgtgtttgtgtgagaataaaaccccttccagactcccaaaaaaataaaaaa 2234
 Db 1795 ttcgtgtttgtgtgagaataaaaccccttccagactcccaaaaaaataaaaaa 1854
 QY 2235 aaaaaa 2241
 Db 1855 aaaaaa 1861

RESULT 4
 US-09-643-893-287
 ; Sequence 287, Application US/09643893
 ; GENERAL INFORMATION:
 ; APPLICANT: Silos-Santiago, Inmaculada
 ; APPLICANT: Dictefano, Peter
 ; APPLICANT: Curtis, Rory A.J.
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES.
 ; FILE OF INVENTION: THEREFOR
 ; FILE REFERENCE: 1600.1161-001
 ; CURRENT APPLICATION NUMBER: US/09/643,893
 ; PRIOR FILING DATE: 2000-08-18
 ; PRIOR APPLICATION NUMBER: 60/150,147
 ; PRIOR FILING DATE: 1999-08-20
 ; NUMBER OF SEQ ID NOS: 412
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 287
 ; LENGTH: 2638
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(2638)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-643-893-287

Query Match 29.0%; Score 649.4; DB 25; Length 2638;
 Best Local Similarity 76.2%; Pred. No. 3.9e-93;
 Matches 830; Conservative 0; Mismatches 246; Indels 13; Gaps 2;

QY 495 gttgaaatcaaaag-aagcaggtggttagtgccagggcagcagccctggaagcagcagcttt 553
 Db 505 gttgaaatcaaaag-aagcaggtggttagtgccagggcagcagccctggaagcagcagcttt 564
 QY 554 ccagggccctctggtcaggaatcctgttgcaagttcccatcatccagagagcagagga 613
 Db 565 gcagaatctctgtcaggaatcctgttgcaagttcccatcatccagagagcagagga 624
 QY 614 ggcctcagctccctcggaaagaaagcctccagcccatcgggtgatttgcagctgaaag 673
 Db 625 tgccctgctgtctctttaaagaagattcccaacccaatgggtgatgcaattgaaag 684
 QY 674 aggcgccagatgctctcatagacaactgtgcgcgaggagagctcaagcgcctccatc 733
 Db 685 gggcacacaaagctatgacagcaatctctagaaacagagaaactaaagcagcctcatt 744
 QY 734 gcttcctcagtcagtcagcagcagcttccctcagtcagcagcagcctcctaaagcagcag 793
 Db 745 ggttcctcagtcagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 804
 QY 794 aacttttagtggaagactctcctcagtcagcagcagcagcagcagcagcagcagcagcagcag 853
 Db 805 tgccttagtgaggagatttttccagcagcagcagcagcagcagcagcagcagcagcagcag 864
 QY 854 taatgagctctcccatcagctcaatggggcgtccttccctcagcctgagcctgctgca 913
 Db 865 tgaggagcctcaccatcagctcaatggggcgtccttccctcagcctgagcctgctgca 924
 QY 914 agggccaccccaaaataactctgtctgggtactgtgactgtcttctccagcgggagcctgtg 973

Query Match				28.9%	Score 647.8	DB 29	Length 2134
Best Local Similarity				76.1%	Pred. No. 7e-93		
Matches 829				Conservative 0	Mismatches 247	Indels 13	Gaps
Qy	495	gttgaatcaaaag-aagcaggtggtagtgctgcaggcggcagccctgaagacgcagcttt	553				
Db	250	gttgaatcaaggtaaagcaggtggtactactacaagttaaaccggagaagaacaactt	309				
Qy	554	ccaggcccccttgctcaggaaactctgttgcaggttcccatcattccaggagcagagga	613				
Db	310	gcagaaattcttctgctcaggaaactctgttgcaggttcccatggtcccaggaaactagagga	369				
Qy	614	ggctccagctgcctccggagaagaagactccagcccctggtgatgttgcagctgaaag	673				
Db	370	tgccctcgtgttctcttaagaaagattccaaaccaatggtgatgccaatgtgaag	429				
Qy	674	agggcccaagctctctgctatagaaactctggcgcgaggagactcaaaagcgtccatct	733				
Db	430	gggcacacaaatgctatgtatgacaaattctagaaagagaactaaaagcactccatt	489				
Qy	734	gcttcctcagtacgatgacagagcagttccctcagtcagactccctaaagccaatgac	793				
Db	490	ggttcctcagatcaagatcaaaataattatctacagtcagatgtccctaaaccaatgac	549				
Qy	794	aactttagtggaagactctgcagtcaccagcgaagttaaattctcatcacacaggtga	853				
Db	550	tgctttagtaggagattttgcagcatcaacaaaattaaattctattacacaacaact	609				
Qy	854	taatggagctctccactcagctgtcaatggggctgcctttccctctggacctgctctgca	913				
Db	610	tgaggagcttaacctcggtagtcacagggtctgctttcccctcggatcaactctcc	669				
Qy	914	agggccaccccaataactctgctgggtactgtgactgtcttcctcagcgggacgtctcg	973				
Db	670	agggaccacaaaataactttgctgggtgactgtgactgtttgcagtggggacttttg	729				
Qy	974	caacagctgca-----gctgcaacaacctgcgcctagctgagcgcgtctcaa	1021				
Db	730	caacaactgcaattgttaataatgttgcaacaactgcatcatgtattgaaagggttaa	789				
Qy	1022	agccataaaggctgtctttgatagaaaactctgaagctttccaccccaaaaattggggaagg	108				
Db	790	agccattaaagcatgtcttgtagaaatccagagctttccagccgcaaaaattgggaagg	849				
Qy	1082	ccgtctggagctgtcaactctgcacacagcaagggtgcaactgtaaagcgtcaggcgtg	114				
Db	850	ccaattgggcaatgtcgaagccccagcacacaagggtgcaactgtcaggaggtcaggctg	909				
Qy	1142	cctgaagaactactgtgagtgtctatgaggccaaaactcatgtttcttccatttgcaaatg	120				
Db	910	cctgaagaattactgcagtgctatgaggcccaaaattatgtttcttctatttgcaaatg	969				
Qy	1202	cattgctgcanaaactatgaagaagatccagaaacgaaaaatgctgatgag					

Qy	1142	cctgagaagactctgtgagtgctatgagcgccaaatcatgtgttcttcatttgcacaatg	1201
Db	910	ccfgaagaattactcgtcgatgctatgagcgccaaattatgtgtctctctatttgcacaatg	969
Qy	1202	cattgtcttcaaaactatgaagaagctccagacgaaaaatcgtgatgagcacaccca	1261
Db	970	cattgtctgcaaaatttttgaagaagccagaaacacattatgcatgagcatgccc	1029
Qy	1262	ctacatggagcctggggactttgagcagcgcattattgtcccgaccgaagttctcag	1321
Db	1030	ctacatgcagctggaggtttggaaggcagccattacactgcacccagcaaaatttcag	1089
Qy	1322	acctcaaaactgagaaaaataggcaggccttctctctgtatctctctgggaagtatga	1381
Db	1090	acttccaagattcagtcacgaaggcggccttccctcatgcatctccctgggaggtgg	1149
Qy	1382	ggccaatgtgctgctgctgctgccaggggtgagaagcagcagcagcagcactgttccc	1441
Db	1150	ggccaatgcgctgctgctgctcagggagagagcccgagaaacacactgctccaa	1209
Qy	1442	aagcttggtgaacagatgatcctggaggagtttggaagggtgctgtcgagatttcca	1501
Db	1210	gtcgtggcagagcagatgatcctggaggaaatttggaagggtgcttaccagatctcca	1269
Qy	1502	catcgattcaaatccaaagggtctgaaattgaatagtcgtgcaagctgtgtaaggggaat	1561

Qy	1142	ctgagagaaacacacgctgagagctgctatgagagggcacaataacacatgctctctccatctgcaaatg	1201
Qy	1143	ctgagagaaacacacgctgagagctgctatgagagggcacaataacacatgctctctccatctgcaaatg	1202
Db	910	cctgagaagattactgcgagtgctaigagggccaaattatgtgtctctattgtgcaaatg	969
Qy	1202	cattgctgtcaaaaactatgaagaaagttccagaaacgaaaaatgctgatgagcacacccca	1261
Qy	1203	ctgagagaaacacacgctgagagctgctatgagagggcacaataacacatgctctctccatctgcaaatg	1262
Db	970	cattggtgtcaaaaattatgaagaaagccagaaacgaaacactaatgagcatgcccacaa	1029
Qy	1262	ctacatgagcctggggacttctgagagcagccattattgtccccagcccaagtctctcaagg	1321
Qy	1263	ctgagagaaacacacgctgagagctgctatgagagggcacaataacacatgctctctccatctgcaaatg	1322
Db	1030	ctacatgcagactggaggtgttggaaggcagccattactctgcacacacgaaattttcaag	1089
Qy	1322	acctcaaaactgaagaaaaatagcagggcctctctctgtatctctctctgggaaagtagtgga	1381
Qy	1323	ctgagagaaacacacgctgagagctgctatgagagggcacaataacacatgctctctccatctgcaaatg	1382
Db	1090	acttccaagattcagtcacga tagcgggccttctctcatgcatctctctcgggaggtggggga	1149
Qy	1382	ggccacatgtgctctgctctgctgctggccagggtaggaaacagacagacagacactgtttcccc	1441
Qy	1383	ctgagagaaacacacgctgagagctgctatgagagggcacaataacacatgctctctccatctgcaaatg	1442
Db	1150	ggccacatgcgctgcctgctgctgctcagggagaagacgagaaacactgctccaa	1209
Qy	1442	aagcttggctgaacagatgatcctgagagagtttggaagggtgctgtcgagattctcca	1501
Qy	1443	ctgagagaaacacacgctgagagctgctatgagagggcacaataacacatgctctctccatctgcaaatg	1502
Db	1210	gtcgtggcagacagatgatcctgagggaaatttggaagggtgcttaccagatctcccca	1269
Qy	1502	cattcgaattcaaatccaaggggctcaaaattgaatagtcgctgcaagctgggtaaagggaat	1561

Db 1270 cactgagtttaactaaggagatgaaatggagtagagataaagtgatgaatcatgtt 1329
Qy 1562 gctgtgagc 1570
Db 1330 gattttgtc 1338

RESULT 6

US-09-471-275-5610/C
; Sequence 5610, Application US/09471275
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: Novel Contigs Obtained
; FILE REFERENCE: 782
; CURRENT APPLICATION NUMBER: US/09/471,275
; CURRENT FILING DATE: 1999-12-23
; EARLIER APPLICATION NUMBER: US 09/235,076
; EARLIER FILING DATE: 1999-01-20
; EARLIER APPLICATION NUMBER: US 09/234,611
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: US 09/240,371
; EARLIER FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: US 09/277,227
; EARLIER FILING DATE: 1999-03-25
; EARLIER APPLICATION NUMBER: US 09/271,490
; EARLIER FILING DATE: 1999-03-18
; EARLIER APPLICATION NUMBER: US 09/293,972
; EARLIER FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: US 09/274,861
; EARLIER FILING DATE: 1999-03-23
; EARLIER APPLICATION NUMBER: US 60/125,453
; EARLIER FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: US 60/126,605
; EARLIER FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: US 09/306,350
; EARLIER FILING DATE: 1999-05-07
; EARLIER APPLICATION NUMBER: US 09/399,720
; EARLIER FILING DATE: 1999-09-21
; EARLIER APPLICATION NUMBER: US 09/404,284
; EARLIER FILING DATE: 1999-09-21
; EARLIER APPLICATION NUMBER:
; EARLIER FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 10451
; SOFTWARE: PLCT_genes Version 1.0
; SEQ ID NO 5610
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1771)...(874)
; OTHER INFORMATION: similar to g14581563 in the genepept database release 114,
; OTHER INFORMATION: Run with FASTX 3.3t00, default parameters
US-09-471-275-5610

Query Match 25.0%; Score 560.4; DB 18; Length 1770;
Best Local Similarity 76.1%; Pred. No. 4.7e-79;
Matches 709; Conservative 0; Mismatches 211; Indels 12; Gaps 1;

Qy 651 atgtgtattgtcagctgaaggagggccagatgctctcatagacaaactgtggcg 710
Db 1770 ATGTGTATGCCAATTTGAAGGGGGCACAAATGCTATATAGACAAATTTAGAAC 1711
Qy 711 agggagctcaagcgctccatctctctcagtcagtcagcagagcagttccctcag 770
Db 1710 AGAAGCTAAAGACACTCCATTTGGTCTCAGTATCAAGATCAAAATAATATCTACAG 1651
Qy 771 tcagagctccctaaagccaatgacaacttagtgggaagactctgccagtcagcgaag 830
Db 1650 TCAGATGTCCTAAACCAATCACTGCTTTAGTAGGGAGATTTTGGCCAGCATCAACAAA 1591

Qy 831 ttaactctcatcacacaggttgataatgagctctcccatcagctgtcactggtgagcgc 890
Db 1590 TTAATCTCATTTACACAACTTGGAGGAGCCTTACCATTGGTAGTCAACGGGTCTGCT 1531
Qy 891 ttccctctgacactgctctgcaaggccaccaccccaataaactctgtctgggtactgtgac 950
Db 1530 TTCCCTCGGGATCAACTCTTCCAGGACCACCAAAAATAAATTTGGTGGGTACTGTGAC 1471
Qy 951 tgcctctccagcgggactctctgcaacagctgca-----gctgcaacacactg 998
Db 1470 TGCTTTGCCAGTGGGACTTTTGAACAACATGCAATTTAATAATTTGTGCAACAACCTG 1411
Qy 999 cgccatgagctcgagcgtctcaaaagccataaagcgtctctgtatagataaactgaaact 1058
Db 1410 CATCATGATATTGAACGGTTTAAAGCCATTAAAGGCTGCTTTGGTAGAAATCCAGAACT 1351
Qy 1059 ttccaacccaaaatgggaaggccgtctgagagctgctaaacttcgacagcgaaggg 1118
Db 1350 TTCCAGCCAAAATTTGGGAAGGCCAATTTGGCAATGTCAAGCCCCAGCACACAAGGG 1291
Qy 1119 tgcactgtgaagcgtcagcgtcgtgaagaactactgtgagtcactgtgagccaaatc 1178
Db 1290 TGCAACTGCAGGAGGTCAAGGCTGCTGAAGAATTTACTGCGAGTGTCTATGAGGCCCAAT 1231
Qy 1179 atgtgtctctccatttgcataatgcatctgtcaaaaactatgaagaagtcagaaacga 1238
Db 1230 ATGTGTTCTTCTATTGCAAAATGCTGTTGCAAAATTTATGAAGAAGCCCAAGACA 1171
Qy 1239 aaaatgctgatgagcacacccactacatggagcctgggactttgagagcagccattat 1298
Db 1170 AAGACACTAATGAGCATGCGCAACTTACATGCAGACTGGAGGTTTGGGAAGGCACCTTAC 1111
Qy 1299 ttgtcccaagccaagttctcagggacctccaaaactgaaaaataagcagcgcctctcc 1358
Db 1110 CTGCCACCAACGAAATTTTCAGGACTTCCAAGATTCAGTACGATAGCGGCTTCTCTCA 1051
Qy 1359 tgtatctctggaagtagtgagggccacatgtgctgcctgctgcccaggtgagaa 1418
Db 1050 TGCACTCTCTGGGAGGTGGTGAGGCCACATGCGGCTGCCCTGCTTGCCTCAGGAGAGAG 991
Qy 1419 gcagagcaggagcactgttcccccaagcttgctgagcagatgatcctggaggagtttga 1478
Db 990 GCCGAGAAAGAACACTGCTCCAAAGTGCCTGGCAGAGCAGATGATCCTGGAGGAATTTGA 931
Qy 1479 aggtgctgtcgcagattctccacatcagttcaagttcaagtcgaaggcgtgaaatgagtag 1538
Db 930 AGGTGCTTTATCACAGATTTCTCCACTGAGTTTAAATCTAAGGGATTGAAAATGGAGTAG 871
Qy 1539 cgtgcaagctgttaagggggaatgcctgtgc 1570
Db 870 AGTATAAGTGTGAATGATGCTGTTGATTTTGTG 839

RESULT 7

US-09-289-768-13239/C
; Sequence 13239, Application US/09289768
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-765
; CURRENT APPLICATION NUMBER: US/09/289,768
; CURRENT FILING DATE: 1999-04-08
; NUMBER OF SEQ ID NOS: 39996
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13239
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(397)

OTHER INFORMATION: n = A,T,C or G
US-09-289-768-13239

Query Match 12.1%; Score 271.8; DB 16; Length 397;
Best Local Similarity 80.5%; Pred. No. 2e-33; 77; Indels 0; Gaps 0;
Matches 318; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1119 tgaactgttaagcgcctcagcgtcctgaagaactactgtgagtgctatgagcccaaatc 1178
DB 396 TGCAACTGCAGGAGGTCAGGCTGCCTGAAGATTACTGCGAGTGCTATGAGGCCCAAAAT 337

QY 1179 atgtgttcttcattgcaaatgcatgttgcataaaactatgaagaagtccagaacga 1238
DB 336 ATGTGTTCTTCTATTGCAAAATGCATTGGTTGCAAAATATTGAAGAAAGCCAGAACGA 277

QY 1239 aaaaactatgagacacccccactacatgagcctggagccttggagccttgcattat 1298
DB 276 AAGACACTAATGAGCATGCAACTACATGCAGACTGGGAGGTTTGAAGCGACCAATTAC 217

QY 1299 ttgtcccccagcaagtctcagagcctccaaactgagaaaaataggcgcctctcc 1358
DB 216 CTGCCACCAACGAATTTTCAGGACTTCCAGATTTCAGTCAGATAGCGGCTTCTCTCA 157

QY 1359 tgtatctctgggaagtagtggagccacatgtgcctgctgctggccaggtgagga 1418
DB 156 TGCATCTCCTGGAGGTGGTGGAGGCCACATGGCCTGCTGCTCAGGGAGAAGAG 97

QY 1419 gcagagcagagcactgttccccaaactggctgagcagatgctcctgagagtttga 1478
DB 96 GCGGAGAAAGAACCCCTGCTCCCAAGTGCCTGGCAGACGATGATCCTGGAGGAATTTGA 37

QY 1479 aggtgctgtcgcagattctccacatcgagttcaa 1513
DB 36 AGTGCTTATCACAGATTCTCCACACTGAGTTAA 2

RESULT 8

US-09-939-397-13239/c
Sequence 13239, Application US/09939397
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 20411-765
CURRENT APPLICATION NUMBER: US/09/939,397
CURRENT FILING DATE: 2001-08-24
PRIOR APPLICATION NUMBER: 09/289,768
PRIOR FILING DATE: 1999-04-08
NUMBER OF SEQ ID NOS: 39996
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 13239
LENGTH: 397
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(397)
OTHER INFORMATION: n = A,T,C or G

US-09-939-397-13239

Query Match 12.1%; Score 271.8; DB 35; Length 397;
Best Local Similarity 80.5%; Pred. No. 2e-33;
Matches 318; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1119 tgaactgttaagcgcctcagcgtcctgaagaactactgtgagtgctatgagcccaaatc 1178
DB 396 TGCAACTGCAGGAGGTCAGGCTGCCTGAAGATTACTGCGAGTGCTATGAGGCCCAAAAT 337

QY 1179 atgtgttcttcattgcaaatgcatgttgcataaaactatgaagaagtccagaacga 1238
DB 336 ATGTGTTCTTCTATTGCAAAATGCATTGGTTGCAAAATATTGAAGAAAGCCAGAACGA 277

QY 1239 aaaatgctgtagagacacccccactacatgagcctggggaactttgagagcagccattat 1298
DB 276 AAGACACTAATGAGCATGCAAACTACATGCAGACTGGAGGTTTGAAGCGCAGCCATTAC 217

QY 1299 ttgtcccccagcaagtctcagagcctccaaactgagaaaaataggcgcctctcc 1358
DB 216 CTGCCACCAACGAATTTTCAGGACTTCCAGATTTCAGTCAGATAGCGGCTTCTCTCA 157

QY 1359 tgtatctctgggaagtagtggagccacatgtgcctgctgctggccaggtgagga 1418
DB 156 TGCATCTCCTGGAGGTGGTGGAGGCCACATGGCCTGCTGCTCAGGGAGAAGAG 97

QY 1419 gcagagcagagcactgttccccaaactggctgagcagatgctcctgagagtttga 1478
DB 96 GCGGAGAAAGAACCCCTGCTCCCAAGTGCCTGGCAGACGATGATCCTGGAGGAATTTGA 37

QY 1479 aggtgctgtcgcagattctccacatcgagttcaa 1513
DB 36 AGTGCTTATCACAGATTCTCCACACTGAGTTAA 2

RESULT 9

PCT-US01-01329-425
Sequence 425, Application PC/TUS0101329
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc., et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA120PCT
CURRENT APPLICATION NUMBER: PCT/US01/01329
CURRENT FILING DATE: 2001-01-17
NUMBER OF SEQ ID NOS: 3506
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 425
LENGTH: 445
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (431)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (438)
OTHER INFORMATION: n equals a,t,g, or c

PCT-US01-01329-425

Query Match 11.1%; Score 248.4; DB 1; Length 445;
Best Local Similarity 73.9%; Pred. No. 1e-29;
Matches 312; Conservative 1; Mismatches 109; Indels 0; Gaps 0;

QY 495 gttgaaatcaagaagcaggtgtagtggcagcgcgagcctgaaacgcagctttc 554
DB 24 gttgaaatcaagaagcaggtgtagtggcagcgcgagcctgaaacgcagctttc 83

QY 555 caggccctctggtcaggaatcctgttgcaagctcccatcctccagagcagagag 614
DB 84 cagaatctcttgcaggaatcctgttgcaagctcccatcctccagagcagagag 143

QY 615 gctccagctgcccctgggaagaagactccagcccatcctggtgatttgcagctgaaagga 674
DB 144 gctccagctgcccctgggaagaagactccagcccatcctggtgatttgcagctgaaagga 203

QY 675 ggcgcccagatgctctgcatagacaactgtggcgaggagagctcaaacgctccatctg 734
DB 204 ggcacacaatgctatgtagacaattctagacacagagaaactaaagcactccattg 263

QY 735 ctctcctcagtaacgatgaccagagcaggttccctcagtcagagcctcctaaagcaagaca 794
DB 264 gttcctcagtaacgatgaccagagcaggttccctcagtcagagcctcctaaagcaagaca 323

QY 795 actttagtggaagacgtctcctcagtcagagcagagtaaaactcctcacacacaggttgat 854

Db 324 gctttagtaggagattttgcccagcatcaacaaaattaaatctcattacacacaactt 383
Qy 855 aatggagcttcccatcagctgtcaatgggctgccttccctctgagccctgctctgcaa 914
Db 384 gaggagccttaccatcggtagtaacagggtctgtcttccctcggtgntcaactntcca 443
Qy 915 gg 916
Db 444 gg 445

RESULT 10

PCT-US01-01339-1301

; Sequence 1301, Application PC/TUS0101339
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01339
; CURRENT FILING DATE: 2001-03-17
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1301
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (431)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (438)
; OTHER INFORMATION: n equals a,t,g, or c

PCT-US01-01339-1301

Query Match 11.1%; Score 248.4; DB 1; Length 445;
Best Local Similarity 73.9%; Pred. No. 1e-29;
Matches 312; Conservative 1; Mismatches 109; Indels 0; Gaps 0;

Qy 495 gttgaatacgaagcagggtgtgtgcccaggcggcagccctgaagacgcagcttcc 554
Db 24 gttgaatacgaagcagggtgtgtgcccaggcggcagccctgaagacgcagcttcc 83
Qy 555 caggccctctgctcaggaatcctgttgcaagttcccatcattcccggaagacgagag 614
Db 84 cagaatctctgctcaggaatcctgttgcaagttcccatcattcccggaagacgagag 143
Qy 615 gcttcagctgcccctggaagaaagactccagcccatggtgattgttcagctgaagaa 674
Db 144 gcttcctgctgttcttgaagaagattcccaaccaaagggtgatacgcaattgaaagg 203
Qy 675 gggccccaagctgctgtcatagacaactgtggcgagggagctcaaaagcgtccatctg 734
Db 204 ggcacacaaatgctatgtatagacaattctagacragagaaactaaaagcactccattg 263
Qy 735 ctctcagtagtacgaccagagcaggttccctcagtcagagctcccttaagccaatgaca 794
Db 264 gtccctcagtagtacaagatcaaaataattatctacagtcagatgtccctaaaccaa 323
Qy 795 actttagtggaagactctcagtagtaccaggaagttaaatctctatcacacaggtgat 854
Db 324 gctttagtaggagatttttccagcatcaacaaaattaaatctcattacacacaactt 383
Qy 855 aatggagcttcccatcagctgtcaatgggctgccttccctctgagccctgctctgcaa 914
Db 384 gaggagccttaccatcggtagtaacagggtctgtcttccctcggtgntcaactntcca 443
Qy 915 gg 916
Db 444 gg 445

RESULT 12

US-10-080-090-425

; Sequence 425, Application US/10080090
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA120C1
; CURRENT APPLICATION NUMBER: US/10/080,090
; CURRENT FILING DATE: 2002-02-22
; Prior Application removed - See File Wrapper or Palm

RESULT 11

US-09-764-891-1301

; Sequence 1301, Application US/09764891
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1301
; LENGTH: 445
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (431)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (438)
; OTHER INFORMATION: n equals a,t,g, or c

US-09-764-891-1301

Query Match 11.1%; Score 248.4; DB 30; Length 445;
Best Local Similarity 73.9%; Pred. No. 1e-29;
Matches 312; Conservative 1; Mismatches 109; Indels 0; Gaps 0;

Qy 495 gttgaatacgaagcagggtgtgtgcccaggcggcagccctgaagacgcagcttcc 554
Db 24 gttgaatacgaagcagggtgtgtgcccaggcggcagccctgaagacgcagcttcc 83
Qy 555 caggccctctgctcaggaatcctgttgcaagttcccatcattcccggaagacgagag 614
Db 84 cagaatctctgctcaggaatcctgttgcaagttcccatcattcccggaagacgagag 143
Qy 615 gcttcagctgcccctggaagaaagactccagcccatggtgattgttcagctgaagaa 674
Db 144 gcttcctgctgttcttgaagaagattcccaaccaaagggtgatacgcaattgaaagg 203
Qy 675 gggccccaagctgctgtcatagacaactgtggcgagggagctcaaaagcgtccatctg 734
Db 204 ggcacacaaatgctatgtatagacaattctagacragagaaactaaaagcactccattg 263
Qy 735 ctctcagtagtacgaccagagcaggttccctcagtcagagctcccttaagccaatgaca 794
Db 264 gtccctcagtagtacaagatcaaaataattatctacagtcagatgtccctaaaccaa 323
Qy 795 actttagtggaagactctcagtagtaccaggaagttaaatctctatcacacaggtgat 854
Db 324 gctttagtaggagatttttccagcatcaacaaaattaaatctcattacacacaactt 383
Qy 855 aatggagcttcccatcagctgtcaatgggctgccttccctctgagccctgctctgcaa 914
Db 384 gaggagccttaccatcggtagtaacagggtctgtcttccctcggtgntcaactntcca 443
Qy 915 gg 916
Db 444 gg 445

RESULT 12

US-10-080-090-425

; Sequence 425, Application US/10080090
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA120C1
; CURRENT APPLICATION NUMBER: US/10/080,090
; CURRENT FILING DATE: 2002-02-22
; Prior Application removed - See File Wrapper or Palm

Query Match 12.5%; Score 279.4; DB 1; Length 486;
 Best Local Similarity 76.4%; Pred. No. 2.4e-39;
 Matches 343; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 546 gcaagcttccagggccctctggtcaggaatcctgttgaagttcccatccacaggag 605
 DB 1 gcaacttgcagaatctcttgcaggaatcctgttgaagttcccatccacaggag 60

QY 606 gcaagagggccctccagctgcccctggaagaagagctccagcccatggtgattgtcag 665
 DB 61 ctagagagtgccctctgctgttctcttaagaagaattccaaacccaatggtgatgccaa 120

QY 666 ctgaaggaggccagagctctctctatagacaactgtggcgaggagctcaaaagcg 725
 DB 121 ttgaaggaggccacacaaatgctatgatagacaattctagaacaagagaactaaaagca 180

QY 726 ctccatctcttccagtagcacgagcagcagcttccctcagtcagagctccctaaag 785
 DB 181 ctccattggttccagtagcacgagcagcagcttccctcagtcagagctccctaaag 240

QY 786 ccaatgacaactttagtggaagactctgcccagtagcacgagcagcagctccctaaag 845
 DB 241 ccaatgagctgtttagtggaagactctgcccagtagcacgagcagcagctccctaaag 300

QY 846 caggttgataatggagctctccatcagctgtcaatggggtgcttccctctggaact 905
 DB 301 caacaacttgaggagccttaccatcggtagtcacacgggtctgttccctcccggaatca 360

QY 906 gctctgcaaggccaccccaaaataactctgtggtgactgtgactgttctccacggcg 965
 DB 421 gactttgcaacaactgcaattgtaataa 449

RESULT 5

US-10-112-699-2370
 ; Sequence 2370, Application US/10112699
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Tongtong
 ; APPLICANT: Wang, Siqing
 ; APPLICANT: Bangur, Alexander
 ; APPLICANT: Gaiger, Alexander
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 ; TITLE OF INVENTION: AND DIAGNOSIS OF CANCER
 ; FILE REFERENCE: 210121.565
 ; CURRENT APPLICATION NUMBER: US/10/112,699
 ; CURRENT FILING DATE: 2002-03-28
 ; NUMBER OF SEQ ID NOS: 2959
 ; SEQ ID NO 2370
 ; LENGTH: 486
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 484
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-10-112-699-2370

Query Match 12.5%; Score 279.4; DB 6; Length 486;
 Best Local Similarity 76.4%; Pred. No. 2.4e-39;
 Matches 343; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 546 gcaagcttccagggccctctggtcaggaatcctgttgaagttcccatccacaggag 605
 DB 1 gcaacttgcagaatctcttgcaggaatcctgttgaagttcccatccacaggag 60

QY 606 gcaagagggccctccagctgcccctggaagaagagctccagcccatggtgattgtcag 665
 DB 61 ctagagagtgccctctgctgttctcttaagaagaattccaaacccaatggtgatgccaa 120

QY 666 ctgaaggaggccagcagagctgtctgtcagtagacaactgtggcgagggagctcaaaagcg 725
 DB 121 ttgaaggaggccacacaaatgctatgatagacaattctagaacaagagaactaaaagca 180

QY 726 ctccatctcttccagtagcacgagcagcagcttccctcagtcagagctccctaaag 785
 DB 181 ctccattggttccctcagtagcacgagcagcagcttccctcagtcagagctccctaaag 240

QY 786 ccaatgacaactttagtggaagactcttccagtagcacgagcagcagctccctaaag 845
 DB 241 ccaatgagctgtttagtggaagactcttccagtagcacgagcagcagctccctaaag 300

QY 846 caggttgataatggagctctccatcagctgtcaatggggtgcttccctctggaact 905
 DB 301 caacaacttgaggagccttaccatcggtagtcacacgggtctgttccctcccggaatca 360

QY 906 gctctgcaaggccaccccaaaataactctgtggtgactgtgactgttctccacggcg 965
 DB 361 actcttcccgagccaccccaaaataactctgtggtgactgtgactgttctccacggcg 420

QY 966 gactttgcaacaactgcaattgtaataa 994
 DB 421 gactttgcaacaactgcaattgtaataa 449

RESULT 6

US-09-721-544-15843
 ; Sequence 15843, Application US/09721544
 ; GENERAL INFORMATION:
 ; APPLICANT: Arterburn, Matthew
 ; APPLICANT: Asghari, Vida
 ; APPLICANT: Damavandi, Simin
 ; APPLICANT: Dickson, Mark
 ; APPLICANT: Drake, Jim
 ; APPLICANT: Drmanac, Radoje
 ; APPLICANT: Engleman, Carrie
 ; APPLICANT: Faulkner, Brandy
 ; APPLICANT: Garcia, Veronica
 ; APPLICANT: Giedt, Gretchen
 ; APPLICANT: Hunter, Kelly
 ; APPLICANT: Jensen, Aaron
 ; APPLICANT: Jones, Lee
 ; APPLICANT: Kita, David
 ; APPLICANT: Labat, Ivan
 ; APPLICANT: Laroya, Mimi
 ; APPLICANT: Lomelli, Michelle
 ; APPLICANT: Nguyen, Phuong
 ; APPLICANT: Nogra, Margie
 ; APPLICANT: Palencia, Servando
 ; APPLICANT: Raisi, Fariba
 ; APPLICANT: Smith, Benjamin
 ; APPLICANT: Tkach, Joe
 ; APPLICANT: Tran, Lien
 ; APPLICANT: Verna, Ron
 ; APPLICANT: Yang, Fei
 ; APPLICANT: Yim, Kenneth
 ; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
 ; TITLE OF INVENTION: From a cDNA Library of Fetal Liver-Spleen
 ; FILE REFERENCE: 728CIP
 ; CURRENT APPLICATION NUMBER: US/09/721,544
 ; CURRENT FILING DATE: 2000-11-21
 ; PRIOR APPLICATION NUMBER: 09/515,128
 ; PRIOR FILING DATE: 1998-02-13
 ; PRIOR APPLICATION NUMBER: 09/034,341
 ; PRIOR FILING DATE: 1998-02-13
 ; NUMBER OF SEQ ID NOS: 24489
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 15843
 ; LENGTH: 952
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 11:10:47 ; Search time 5416.12 Seconds
(without alignments)
8245.244 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 2134
Sequence: 1 aattcgggtcgaagcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_em.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_un.*
- 28: em_vi.*
- 29: em_hum.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description

1	2127.6	99.7	2134	9	HSU86074	
c	2	915	42.9	157269	2	AP003096
3	672.2	31.5	2215	10	AB057423	
4	662	31.0	1861	10	MMU77383	
5	647.8	30.4	2241	10	MMU67176	
6	208.6	9.8	157269	2	AP003096	
7	152.8	7.2	188859	2	AP003732	
8	139.6	6.5	22856	10	AB057422	
c	9	122.4	5.7	188859	2	AP003732
10	122.2	5.7	17953	2	AC097959	
11	108.4	5.1	40893	3	CBRG45E19	
12	102	4.8	39553	3	CEJCB	
13	102	4.8	330724	2	CEY67H2	
c	14	87.4	4.1	179953	2	AC097959
15	83	3.9	214055	2	AC099385	
c	16	83	3.9	247253	2	AC103272
17	83	3.9	297288	2	AC108533	
c	18	76.4	3.6	94362	8	AC005405
c	19	76.4	3.6	170020	8	AP003256
c	20	76.4	3.6	180206	2	AP003274
21	74.6	3.5	187360	9	AC021105	
22	73.6	3.4	97714	8	ATF25024	
23	73.6	3.4	101679	8	ATF19B15	
24	73.6	3.4	198429	8	ATCHRIV70	
25	71.2	3.3	100047	8	AC006081	
26	69.8	3.3	64971	2	AC107744	
c	27	69.6	3.3	128463	2	AP003619
c	28	69.6	3.3	159947	2	AP003763
29	69.2	3.2	27555	2	AC012903	
c	30	69.2	3.2	146153	3	AC007452
c	31	69.2	3.2	159065	3	AC091634
32	69.2	3.2	264646	3	AE003818	
33	65	3.0	188844	2	AC099384	
c	34	57.4	2.7	157285	9	AC068483
35	56.8	2.7	1793	8	AY046019	
36	56.8	2.7	2298	8	AF204059	
37	56.8	2.7	2431	8	AF206324	
38	55.2	2.6	2146	8	AF205142	
c	39	55	2.6	188844	2	AC099384
40	53.2	2.5	2153	9	AB049880	
41	53.2	2.5	2296	9	HSM802712	
c	42	53.2	2.5	125020	9	AF429315
43	52.8	2.5	13684	3	AE001403	
44	50.8	2.4	253305	3	PFMAL3P7	
c	45	50.4	2.4	368	6	AX284242

ALIGNMENTS

RESULT	1	HSU86074	2134 bp	mRNA	linear	PRI 13-APR-1999
LOCUS	HSU86074	Homo sapiens tesmin mRNA, complete cds.				
DEFINITION	U86074					
ACCESSION	U86074.1	GI:4581562				
VERSION						
KEYWORDS						
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS		1 (bases 1 to 2134)				
TITLE		Sugihara,T., Wadhwa,R., Kaul,S.C. and Mitsui,Y.				
JOURNAL		A novel testis-specific metallothionein-like protein, tesmin, is an early marker of male germ cell differentiation				
MEDLINE		Genomics 57 (1), 130-136 (1999)				
REFERENCE		99208669				
AUTHORS		2 (bases 1 to 2134)				
TITLE		Sugihara,T.				
JOURNAL		Direct Submission				
TITLE		Submitted (20-JAN-1997) NIBH, AIST, Higashil-1, Tsukuba, Ibaraki				
FEATURES		305, Japan				
		Location/Qualifiers				

[illegible]

1467

5

D	b	130535	GGCTCCCTGCATCCCTGGGCCACAGGGGAGTTTACTCAGNGCTCTCTGAAGATGTGGCAA	130476
Q	y	1468	cccatgccccctttctgaggaggtgcattggcctgagcatttgttgtctggtgccagagga	1527
D	b	130475	CCCATTGCCCTTTTCTTGAGGAGGTGCATGGCTGAGCAATTGTTGTCGTGCCCAGAGA	130416
Q	y	1528	gaagaacttggttcccatagtcctggagaagtgtctcgaggcggc-----	1573
D	b	130415	GAGAGCTTGGGTTCCCATAGTCTCTGGAGAGTGTCTGCAGGGCGGCGGAGGCAGAGCNG	130356
Q	y	1574	-----ggaggccagagca	1586
D	b	130355	GGCAGGGGAGGGCACAGCAGGCCCTTCGAAGGGCAGAGGGGGGCGAGGGGAGGCAGAGCA	130296
Q	y	1587	ggccctcgagagagacctaactctgttcgactcttcctctcagagaaatgttgcctctgaggc	1646
D	b	130295	GGCCCTCGGAGAGCTCACACTCTGGTGCACATCTTCCTCTCAGAGAAATGTTGCTCTGGAGGC	130236
Q	y	1647	tgcctcgtcatgaaaacctaatggttcttctgttgttttccaattatttagaaaataagt	1706
D	b	130235	TGCTCTGCATGA AACCTAATGGTTCTTGTTGTTTTTCAAAATATTTAGAAAAATAAGT	130176
Q	y	1707	tctccggatgggctgtgtgatcacaccactaaaaactctctagagaaactactgaacaacctaaa	1766
D	b	130175	TC TCCGGATGGGTGTTTGATACC ACTTA AAAATCTCTAGAGAACCTACCTGAAACCTAAA	130116
Q	y	1767	gattttctgtagcgtagatatattcccccagagacacgcgaactgtaacttttctctaaaggc	1826
D	b	130115	GATTTTCTGTAGCGTAGATATTTC C C C A G A G G C A C G C G A A C T G T C A G T C T T C C T A A G G C	130056
Q	y	1827	ccccggagacgcaggcaatkggggcctcgcaggccaggccttgccaccagcatgtctctgagt	1886
D	b	130055	CCCCGGAGACGCAGGCAATGGGGGCTCCGAGGCCAGGCTTGACACAGCATGCTCTGAGT	129996
Q	y	1887	tagaggactaaaattatccagttctctgtgtttctacttgaaattggtgaaaaagctct	1946
D	b	129995	TAGAGGACTTAAATPATCAGTTTCTCTGTGTTTTCTACTTGAA TTGGGAAAAGCTCT	129936
Q	y	1947	attatccaattaaactctccataattattgttgtgaatatattattgtttgttaaacatgt	2006
D	b	129935	ATTATCCAATTAACTTCTCCATAATTATTGTTGTTGTAATATTATTATTGTTGTAAACAATG	129876

RESULT	5				
MMU67176					
LOCUS					
DEFINITION	MMU67176	2241 bp	mRNA	linear	ROD 13-APR-1999
ACCESSION	U67176	Mus musculus tesmin-1 mRNA, complete cds.			
VERSION	U67176.1	Gr:4581558			

Web site: <http://hap.gsc.riken.go.jp/>
 Contact: hattori@gsc.riken.go.jp
 ----- Project Information
 Center project name: HumDraft11
 Center clone name: CTD-2007L18
 ----- Summary Statistics

Sequencing vector: PCR products; 100% of reads
 Chemistry: Dye-terminator ET-amersham; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 147566 bases at least Q40
 Consensus quality: 151779 bases at least Q30
 Consensus quality: 153479 bases at least Q20
 Insert size: 154169; sum-of-contigs
 Quality coverage: 8.32x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

```

1 11499 contig of 11499 bp in length
11600 13042 contig of 7443 bp in length
19143 26427 contig of 7285 bp in length
26528 38740 contig of 12213 bp in length
38841 55483 contig of 16643 bp in length
55584 67548 contig of 11965 bp in length
74728 contig of 7080 bp in length
79676 88957 contig of 4747 bp in length
89058 93854 contig of 9282 bp in length
93955 98914 contig of 4960 bp in length
102774 105297 contig of 3659 bp in length
105398 110377 contig of 4980 bp in length
110478 117540 contig of 7063 bp in length
117641 120176 contig of 2536 bp in length
120277 125707 contig of 5431 bp in length
125808 128247 contig of 2440 bp in length
128348 131115 contig of 2768 bp in length
131216 134071 contig of 2856 bp in length
134172 136288 contig of 2117 bp in length
136289 136388 contig of 100 bp
136389 139198 contig of 2810 bp in length
139199 139298 contig of 100 bp
139299 141579 contig of 2281 bp in length
141580 141679 contig of 100 bp
141680 143825 contig of 2146 bp in length
143826 143925 contig of 100 bp
143926 146191 contig of 2266 bp in length
146192 146291 contig of 100 bp
146292 148382 contig of 2091 bp in length
148383 148482 contig of 100 bp
148483 149666 contig of 1184 bp in length
149667 149766 contig of 100 bp
149767 151747 contig of 1981 bp in length
151748 151847 contig of 100 bp
151848 153751 contig of 1904 bp in length
153752 153851 contig of 1026 bp in length
153852 154877 contig of 100 bp
154878 154977 contig of 100 bp
154978 156158 contig of 1181 bp in length
156159 156258 contig of 100 bp
156259 157269 contig of 1011 bp in length.
156259 157269 contig of 1011 bp in length.

```

* NOTE: This is a 'working draft' sequence. It currently consists of 32 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 11499: contig of 11499 bp in length
11500 11599: gap of 100 bp
11600 19042: contig of 7443 bp in length
19043 19142: gap of 100 bp
19143 26427: contig of 7285 bp in length
26428 26527: gap of 100 bp
26528 38740: contig of 12213 bp in length
38741 38840: gap of 100 bp
38841 55483: contig of 16643 bp in length
55484 55583: gap of 100 bp
55584 67548: contig of 11965 bp in length

```

```

* 67549 67648: gap of 100 bp
* 67649 74728: contig of 7080 bp in length
* 74729 74828: gap of 100 bp
* 74829 79575: contig of 4747 bp in length
* 79576 79675: gap of 100 bp
* 79676 88957: contig of 9282 bp in length
* 88958 89057: gap of 100 bp
* 89058 93854: contig of 4797 bp in length
* 93855 93954: gap of 100 bp
* 93955 98914: contig of 4960 bp in length
* 98915 99014: gap of 100 bp
* 99015 102673: contig of 3659 bp in length
* 102674 102773: gap of 100 bp
* 102774 105297: contig of 2524 bp in length
* 105298 105397: gap of 100 bp
* 105398 110377: contig of 4980 bp in length
* 110378 110477: gap of 100 bp
* 110478 117540: contig of 7063 bp in length
* 117541 117640: gap of 100 bp
* 117641 120176: contig of 2536 bp in length
* 120177 120276: gap of 100 bp
* 120277 125707: contig of 5431 bp in length
* 125708 125807: gap of 100 bp
* 125808 128247: contig of 2440 bp in length
* 128248 128347: gap of 100 bp
* 128348 131115: contig of 2768 bp in length
* 131116 131215: gap of 100 bp
* 131216 134071: contig of 2856 bp in length
* 134072 134171: gap of 100 bp
* 134172 136288: contig of 2117 bp in length
* 136289 136388: gap of 100 bp
* 136389 139198: contig of 2810 bp in length
* 139199 139298: gap of 100 bp
* 139299 141579: contig of 2281 bp in length
* 141580 141679: gap of 100 bp
* 141680 143825: contig of 2146 bp in length
* 143826 143925: gap of 100 bp
* 143926 146191: contig of 2266 bp in length
* 146192 146291: gap of 100 bp
* 146292 148382: contig of 2091 bp in length
* 148383 148482: gap of 100 bp
* 148483 149666: contig of 1184 bp in length
* 149667 149766: gap of 100 bp
* 149767 151747: contig of 1981 bp in length
* 151748 151847: gap of 100 bp
* 151848 153751: contig of 1904 bp in length
* 153752 153851: gap of 100 bp
* 153852 154877: contig of 1026 bp in length
* 154878 154977: gap of 100 bp
* 154978 156158: contig of 1181 bp in length
* 156159 156258: gap of 100 bp
* 156259 157269: contig of 1011 bp in length.

```

FEATURES

```

source
1..157269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="CTD-2007L18"
/note="assembly_fragment"
11600..19042
/note="assembly_fragment"
19143..26427
/note="assembly_fragment"
26528..38740
/note="assembly_fragment"
38841..55483
/note="assembly_fragment"
55584..67548
/note="assembly_fragment"
67649..74728
/note="assembly_fragment"

```

```

misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```



```

* 1 19831: contig of 19831 bp in length
* 19832 19931: gap of 100 bp
* 19932 31497: contig of 11566 bp in length
* 31498 31597: gap of 100 bp
* 31598 44647: contig of 13050 bp in length
* 44648 44747: gap of 100 bp
* 44748 54198: contig of 9451 bp in length
* 54199 54298: gap of 100 bp
* 54299 63363: contig of 9065 bp in length
* 63364 63463: gap of 100 bp
* 63464 71678: contig of 8215 bp in length
* 71679 71778: gap of 100 bp
* 71779 79960: contig of 8182 bp in length
* 79961 80060: gap of 100 bp
* 80061 87061: contig of 7001 bp in length
* 87062 87161: gap of 100 bp
* 87162 93934: contig of 6773 bp in length
* 93935 94034: gap of 100 bp
* 94035 101099: contig of 7065 bp in length
* 101100 101199: gap of 100 bp
* 101200 108623: contig of 7424 bp in length
* 108624 108723: gap of 100 bp
* 108724 114505: contig of 5782 bp in length
* 114506 114605: gap of 100 bp
* 114606 119720: contig of 5115 bp in length
* 119721 119820: gap of 100 bp
* 119821 124688: contig of 4868 bp in length
* 124689 124788: gap of 100 bp
* 124789 130420: contig of 5632 bp in length
* 130421 130520: gap of 100 bp
* 130521 134923: contig of 4403 bp in length
* 134924 135023: gap of 100 bp
* 135024 138158: contig of 3135 bp in length
* 138159 138258: gap of 100 bp
* 138259 143086: contig of 4828 bp in length
* 143087 143186: gap of 100 bp
* 143187 146616: contig of 3430 bp in length
* 146617 146716: gap of 100 bp
* 146717 150135: contig of 3419 bp in length
* 150136 150235: gap of 100 bp
* 150236 153465: contig of 3230 bp in length
* 153466 153565: gap of 100 bp
* 153566 156263: contig of 2698 bp in length
* 156264 156363: gap of 100 bp
* 156364 158882: contig of 2519 bp in length
* 158883 158982: gap of 100 bp
* 158983 161109: contig of 2127 bp in length
* 161110 161209: gap of 100 bp
* 161210 164035: contig of 2826 bp in length
* 164036 164135: gap of 100 bp
* 164136 166612: contig of 2477 bp in length
* 166613 166712: gap of 100 bp
* 166713 168665: contig of 1953 bp in length
* 168666 168765: gap of 100 bp
* 168766 171614: contig of 2849 bp in length
* 171615 171714: gap of 100 bp
* 171715 173550: contig of 1836 bp in length
* 173551 173650: gap of 100 bp
* 173651 174290: contig of 640 bp in length
* 174291 174390: gap of 100 bp
* 174391 176225: contig of 1835 bp in length
* 176226 176325: gap of 100 bp
* 176326 178179: contig of 1854 bp in length
* 178180 178279: gap of 100 bp
* 178280 179640: contig of 1361 bp in length
* 179641 179740: gap of 100 bp
* 179741 181689: contig of 1949 bp in length
* 181690 181789: gap of 100 bp
* 181790 183060: contig of 1271 bp in length
* 183061 183160: gap of 100 bp
* 183161 184208: contig of 1048 bp in length
* 184209 184308: gap of 100 bp
* 184309 185664: contig of 1356 bp in length

```

```

* 185665 185764: gap of 100 bp
* 185765 187616: contig of 1852 bp in length
* 187617 187716: gap of 100 bp
* 187717 188859: contig of 1143 bp in length.
FEATURES
      Location/Qualifiers
          1..188859
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="11"
             /map="11q"
             /clone="RP11-682D22"
          1..19831
             /note="assembly_fragment"
          19932..31497
             /note="assembly_fragment clone_end.SP6 vector_side:left"
          31598..44647
             /note="assembly_fragment"
          44748..54198
             /note="assembly_fragment"
          54299..63363
             /note="assembly_fragment"
          63464..71678
             /note="assembly_fragment"
          71779..79960
             /note="assembly_fragment"
          80061..87061
             /note="assembly_fragment"
          87162..93934
             /note="assembly_fragment"
          94035..101099
             /note="assembly_fragment"
          101200..108623
             /note="assembly_fragment"
          108724..114505
             /note="assembly_fragment"
          114606..119720
             /note="assembly_fragment"
Query Match
Best Local Similarity 7.2%; Score 152.8; DB 2; Length 188859;
Matches 166; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 239 gcatgacccagttgaatcaaggtgaagcaggtggtactactacaagtaataatccgaa 298
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141012 GCTTTTCTAGGTTGAATCAAGGAAGCAGGTTGTTACTACTAAGTAATAATCCGAA 141071
QY 299 gaagcaacttgcagaatctcttgcagggaatctctggtgcaagttccatggtccag 358
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141072 GAAGCAACTTTCAGAAATCTTCTGCTCAGGAATCTGTTGCAAGTCCCATGTCCTCCAG 141131
QY 359 gaactagagatgccctcctgctcttcttcttaagaagattcccaaccaatggtgatgc 418
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141132 GAACTAGAGATGCCCTCCTGCTGTTCTCTTTAGAAAGATTCCCAACCATGTAATTATG 141191
QY 419 caattgaa 426
      ||| |||
Db 141192 CTAAGAA 141199
RESULT 8
AB057422
LOCUS AB057422 Mus musculus gene for tesmin, complete cds, strain:129.
DEFINITION AB057422
ACCESSION AB057422.1 GI:15617419
VERSION AB057422.1
KEYWORDS
SOURCE Mus musculus (strain:129) DNA.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Sutou,S., Miwa,K., Matsuura,T., Kawasaki,Y., Ohinata,Y. and Mitsui,Y.
TITLE Structure of the tesmin gene encoding a testis-specific persistent

```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE AUTHORS

1 (bases 1 to 179953)
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alsbrooks, S.L., Amarantunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Bimaga, M., Brown, E., Brown, M., Bryant, N.P., Bouch, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dedrich, D.A., Delaney, K.R., Delgado, O., Dunn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseg, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, M., L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Stinson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczkyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 179953)
 Worley, K.C.

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:16901822.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GFSK
 Center clone name: CH230-114G10
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329First call to findPhrapList
 Consensus quality: 163172 bases at least Q40
 Consensus quality: 168807 bases at least Q30
 Consensus quality: 174078 bases at least Q20
 Estimated insert size: 160479; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/GenBank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

COMMENT

* consists of 62 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 6639: contig of 6639 bp in length
 6739: gap of unknown length
 12683: contig of 5944 bp in length
 12684: gap of unknown length
 12783: contig of 4084 bp in length
 16867: gap of unknown length
 16868: contig of 5962 bp in length
 22929: gap of unknown length
 23029: gap of unknown length
 30265: contig of 7236 bp in length
 30365: gap of unknown length
 36366: contig of 6024 bp in length
 36390: gap of unknown length
 42068: contig of 5579 bp in length
 42168: gap of unknown length
 46728: contig of 4560 bp in length
 46828: gap of unknown length
 52208: contig of 5380 bp in length
 52308: gap of unknown length
 56609: contig of 4301 bp in length
 56709: gap of unknown length
 59548: contig of 2839 bp in length
 59648: gap of unknown length
 62710: contig of 3062 bp in length
 62810: gap of unknown length
 66227: contig of 3417 bp in length
 66327: gap of unknown length
 69089: contig of 2762 bp in length
 69189: gap of unknown length
 72645: contig of 3456 bp in length
 72745: gap of unknown length
 76817: contig of 4072 bp in length
 76917: gap of unknown length
 80040: contig of 3123 bp in length
 80140: gap of unknown length
 83579: contig of 3439 bp in length
 83679: gap of unknown length
 86465: contig of 2786 bp in length
 86465: gap of unknown length
 90515: contig of 3950 bp in length
 90615: gap of unknown length
 94222: contig of 3607 bp in length
 94322: gap of unknown length
 97924: contig of 3602 bp in length
 98024: gap of unknown length
 99751: contig of 1727 bp in length
 99851: gap of unknown length
 102642: contig of 2791 bp in length
 102742: gap of unknown length
 106798: contig of 4056 bp in length
 106898: gap of unknown length
 110343: contig of 3445 bp in length
 110443: gap of unknown length
 112934: contig of 2491 bp in length
 113034: gap of unknown length
 116034: contig of 3000 bp in length
 116134: gap of unknown length
 119554: contig of 3420 bp in length
 119654: gap of unknown length
 12029: contig of 2375 bp in length
 122129: gap of unknown length
 125046: contig of 2917 bp in length
 125146: gap of unknown length
 127444: contig of 2298 bp in length
 127544: gap of unknown length
 129251: contig of 1707 bp in length
 129351: gap of unknown length

```

* 129352 131551: contig of 2200 bp in length
* 131551 gap of unknown length
* 131552 133716: contig of 2065 bp in length
* 133716 gap of unknown length
* 133717 136695: contig of 2879 bp in length
* 136695 gap of unknown length
* 136696 138444: contig of 1649 bp in length
* 138444 gap of unknown length
* 138445 140364: contig of 1820 bp in length
* 140364 gap of unknown length
* 140365 140464: gap of unknown length
* 140465 142056: contig of 1592 bp in length
* 142057 142156: gap of unknown length
* 142157 144300: contig of 2144 bp in length
* 144301 144400: gap of unknown length
* 144401 145630: contig of 1230 bp in length
* 145631 145730: gap of unknown length
* 145731 147565: contig of 1835 bp in length
* 147566 147665: gap of unknown length
* 147666 149525: contig of 1860 bp in length
* 149526 149625: gap of unknown length
* 149626 151295: contig of 1670 bp in length
* 151296 151395: gap of unknown length
* 151396 152702: contig of 1307 bp in length
* 152703 152802: gap of unknown length
* 152803 154292: contig of 1490 bp in length
* 154293 154392: gap of unknown length
* 154393 155990: contig of 1598 bp in length
* 155991 156090: gap of unknown length
* 156091 157685: contig of 1595 bp in length
* 157686 157785: gap of unknown length
* 157786 159784: contig of 1999 bp in length
* 159785 159884: gap of unknown length
* 159885 161347: contig of 1463 bp in length
* 161348 161447: gap of unknown length
* 161448 162678: contig of 1231 bp in length
* 162679 162778: gap of unknown length
* 162779 163996: contig of 1218 bp in length
* 163997 164096: gap of unknown length
* 164097 166978: contig of 2882 bp in length
* 166979 167078: gap of unknown length
* 167079 168859: contig of 1781 bp in length
* 168860 168959: gap of unknown length
* 168960 170817: contig of 1858 bp in length
* 170818 170917: gap of unknown length
* 170918 172018: contig of 1101 bp in length
* 172019 172118: gap of unknown length

Query Match 5.7%; Score 122.2; DB 2; Length 179953;
Best Local Similarity 71.7%; Pred. No. 1.6e-15;
Matches 160; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 892 ctgcaggagctcagctgctgaggaattactcgcgagtgctagagcccaaaattatgtg 951
Db 58954 CTGTGGGGGGAAGCCTGATTTTCATCAAGCATTATTTTATCATCAGCGCCAAATACAGTG 59013

QY 952 ttctctatttgcacattgctggttgcacaaattatgaagaagcccgagaaagac 1011
Db 59014 TTCTTCATTGCAATGCTTCTGCAAAATCTATGAAGAAGCCGAGAGCGAAAAAT 59073

QY 1012 actaatgagcatgccaactacatcagactgaggttttgaaggcagccattacgtgc 1071
Db 59074 GCTGATGAGCAGCCCGCTTACATGAGCGCTGGGAGCTTTGAGAGCAGCCACCATCTGTC 59133

QY 1072 accaagaaatttccaggacttccaagattcagtcagatag 1114
Db 59134 CCCAGCCAAATTCAGGAGCTCCAGACTGAGAAAAATAGG 59176

RESULT 11
CBRG45E19 40893 bp DNA linear INV 04-NOV-2000
LOCUS CBRG45E19
DEFINITION Caenorhabditis briggsae cosmid G45E19, complete sequence.
ACCESSION AC084631

```

```

VERSION AC084631.1 GI:11095080
KEYWORDS HTG.
SOURCE Caenorhabditis briggsae.
ORGANISM Caenorhabditis briggsae
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 40893)
AUTHORS Washington University Genome Sequencing Center.
TITLE The C. briggsae Genome Sequencing Project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 40893)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: jsplie@watson.wustl.edu

FEATURES
source 1..40893
/organism="Caenorhabditis briggsae"
/strain="GujArat G16"
/db_xref="taxon:6238"
/clone="G45E19"

BASE COUNT 13086 a 7791 c 7927 g 12089 t
ORIGIN

Query Match 5.1%; Score 108.4; DB 3; Length 40893;
Best Local Similarity 62.7%; Pred. No. 1.1e-12;
Matches 188; Conservative 0; Mismatches 106; Indels 6; Gaps 1;

QY 698 tactgtgactgcttccagtgaggactttgcaacaactgaattgtaattgtgc 757
Db 38000 TACTGTGACTGCTTCGCGAAGGAGAGAGTTTGGCGTGACTGCAACTGCAGGAGCTCCAC 38059

QY 758 acaacttgcatcatgatattgaacggtttaagccattaagcagtcgtcttggtagaaat 817
Db 38060 AACAAATATAGATTATGACAGCCCAACGATCAAAAGCCATTTCGTTCAGTCTTTGGAGCGAAAT 38119

QY 818 ccagaagcttccagcccaaaattgggaaggcccaattgggca-----atgtcaagccc 871
Db 38120 CCAAAATGCTTTCAAGCGGAAAAATCGGAATAGCTCGTGGTGAACCCGCGGATATTGAACGC 38179

QY 872 cagcaacaagaagggtgcaactgcagaggtcaggtcgtcgaagaattactcagagtcg 931
Db 38180 TTGCATCAAAAGGGATGCGCCTGCAAGAAAAGCGGTTCCTGAAAACATATTTCGAGAGTC 38239

QY 932 tatgagcccaaaattatgtgttcttcttatttgcacaaatgcattggttgcacaaatgatga 991
Db 38240 TAGGAAGCAAAAGTGCCTGTACTGATCGATGCAAGTGAAGGCTGTGAGAAATACGGAA 38299

RESULT 12
CEJCB8
LOCUS Caenorhabditis elegans cosmid JCB8, complete sequence.
DEFINITION Caenorhabditis elegans.
ACCESSION Z82274
VERSION Z82274.1 GI:2814248
KEYWORDS HTG.
ORGANISM Caenorhabditis elegans.
SOURCE Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (sites)

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

AUTHORS

Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium Science 282 (5396), 2012-2018 (1998)

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

note.

Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium Science 282 (5396), 2012-2018 (1998)

99069613

The *C. elegans* Sequencing Consortium.

2 (bases 1 to 39553)

Lightning, J.

Direct Submission

Submitted (11-NOV-1996) Nematode Sequencing Project, Sanger

Institute, Hinxton, Cambridge CB10 1SA, England and Department of

Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:

jes@sanger.ac.uk or rwenematode.wustl.edu

On Jan 28, 1998 this sequence version replaced gi:2558532.

Coding regions below are predicted from computer analysis, using

predictions from GeneFinder (P. Green, U. Washington), and other

available information.

Current sequence finishing criteria for the *C. elegans* genome

sequencing consortium are that all bases are either sequenced

unambiguously on both strands, or on a single strand with both a

dye primer and dye terminator reaction, from distinct subclones.

Exceptions are indicated by an explicit note.

This sequence is the entire insert of clone JC8.

The true right end of clone C48D1 is at 6889 in this sequence. The

start of this sequence (1..6889) overlaps with the end of sequence

Z81049.

The end of this sequence (39450..39553) overlaps with the start of

sequence AL32951.

For a graphical representation of this sequence and its analysis

see: - [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=JC8)

name=JC8

IMPORTANT: This sequence is NOT necessarily the entire insert of

the specified clone. It may be shorter because we only sequence

overlapping sections once, or longer because we arrange for a small

overlap between neighbouring submissions.

Location/Qualifiers

1..39553

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="IV"

/clone="JC8"

join(4308..4352,4737..4953,5892..5902)

/gene="JC8.11"

join(4308..4352,4737..4953,5892..5902)

/gene="JC8.11"

/note="cDNA EST yk245a7.3 comes from this gene"

cDNA EST yk245a7.5 comes from this gene"

/codon_start=1

/protein_id="CAB05233.2"

/db_xref="GI:14530485"

/db_xref="SPTREMBL:O62296"

/translation="MGSWNYSIQGTWGEVDENLQKHEDSLRYETASNDQLQKIRDDSD

YFKAMGEIDPIELPFRDRDSSIPACQVINTVESQPIRHLINAD"

complement(join(7174..7377,8392..9150,10178..10645))

/gene="JC8.2"

complement(join(7174..7377,8392..9150,10178..10645))

/gene="JC8.2"

/note="contains similarity to Pfam domain: PF00400 (WD

domain, G-beta repeat), Score=44.5, E-value=7.6e-10, N=2

cDNA EST yk329e7.3 comes from this gene

cDNA EST yk404a8.3 comes from this gene

cDNA EST yk329e7.5 comes from this gene

cDNA EST yk404a8.5 comes from this gene"

/codon_start=1

/protein_id="CAB05230.2"

/db_xref="GI:6425181"

/db_xref="SPTREMBL:Q9NAQ2"

/translation="MKTKPELMDGLVMSVLSIQRGVAKEIPDKIKLDEKLKEL

IEGAIPSTONETELDDDESDNEKLGRLRGDVEKMDSEFEKKYMKGYNEIKD

EPGQCGMGAMTYSKDDPYVTEQVDSDEEEKIMVKDDNNMVAVAKIDKGFET

LECVYNEADSWYCHHDIYILDAPPLCIPEQVDPGNEETGKGNLLAVGTMNSIHTW

DLOITMTATFTLTGKKERVKGAARKRONSAQCHTDAVISLAWNRITTHVLSGGA

DKTVVLWDLDEAKPAQIIPQGGGIEQTGMKWHNPNESTFLLLTGMKGQVNVVDCRESSGN

gene

CDS

ASAAKFDGQIEKVIWVHNPFTAFCSDDGRRLRLHLDMRKPGCELGWEGVADHGPIGL
TISALTKGLVTVGEDSMNVKVEDTNGGIEKVHSEKLTIGELHCAQFNPVNAVLS
VGTAAADLIRIIDLTKPEPVKAFSE"
complement(join(11334..11829,12629..12735,12794..12924,
13189..13450))
/gene="JC8.13"
complement(join(11334..11829,12629..12735,12794..12924,
13189..13450))
/gene="JC8.13"
/codon_start=1
/protein_id="CAD21646.1"
/db_xref="GI:18369715"
/translation="MASCEMCPVAEINFRSRNHLIDQLEATSTSSGPRSKYVADPT
MKVEYSRSAAATHKYKPELLRPVLOHTIDYLLDYSPEKDRRAITSKOPASIF
SFVSDRLRSIQDMIMQNLRCDFTVILLIEKMLPYIETDCACKMAKQCSYDPKLHDFQ
LESCFGRWFEVNSNDIIPNLISACFFERQLHRKSLIHLDFLRKKPLFHSFNLI
QSISSFSYNNRRFFDIFQOOLPLKHSISDCVSLIRSAMKNISVAFPTPVARLPS
QLLSDMIGFANLEFFDFLRYSVIDEQGNILISALRPVVEISYQDFSSROYE"
complement(join(13717..13814,13864..13971,14182..14436,
14562..14598))
/gene="JC8.3a"
complement(join(13717..13814,13864..13971,14182..14436,
14562..14598))
/gene="JC8.3a"
/note="predicted using GeneFinder
contains similarity to Pfam domain: PF00298 (Ribosomal
protein L11), Score=176.7, E-value=1.2e-49, N=1
cDNA EST EMBL:T00775 comes from this gene; cDNA EST
yk30d2.5 comes from this gene; cDNA EST yk139d10.3 comes
from this gene; cDNA EST yk139d10.5 comes from this gene
cDNA EST yk122f7.5 comes from this gene; cDNA EST
yk149h7.5 comes from this gene; cDNA EST yk151h12.5 comes
from this gene; cDNA EST yk165e9.5 comes from this gene
cDNA EST yk149h7.3 comes from this gene; cDNA EST
yk213e10.3 comes from this gene; cDNA EST yk236f7.3 comes
from this gene; cDNA EST yk241b1.3 comes from this gene
cDNA EST yk254h9.3 comes from this gene; cDNA EST
yk268g12.3 comes from this gene; cDNA EST yk269g3.3 comes
from this gene; cDNA EST yk273e7.3 comes from this gene
cDNA EST yk278d11.3 comes from this gene; cDNA EST
yk284f2.3 comes from this gene; cDNA EST yk335c10.3 comes
from this gene; cDNA EST yk335g12.3 comes from this gene
cDNA EST yk299a2.3 comes from this gene; cDNA EST
yk372f10.3 comes from this gene; cDNA EST yk391h1.3 comes
from this gene; cDNA EST yk374g7.3 comes from this gene
cDNA EST yk376a3.3 comes from this gene; cDNA EST
yk402b6.3 comes from this gene; cDNA EST yk417g2.3 comes
from this gene; cDNA EST yk425f6.3 comes from this gene
cDNA EST yk448e7.3 comes from this gene; cDNA EST
yk452b3.3 comes from this gene; cDNA EST yk452g7.3 comes
from this gene; cDNA EST yk474a2.3 comes from this gene
cDNA EST yk464d9.3 comes from this gene; cDNA EST
yk409h4.3 comes from this gene; cDNA EST yk485e8.3 comes
from this gene; cDNA EST yk485e8.3 comes from this gene
cDNA EST yk446g3.3 comes from this gene; cDNA EST
yk475e10.3 comes from this gene; cDNA EST yk483g3.3 comes
from this gene; cDNA EST yk501g7.3 comes from this gene
cDNA EST yk505g10.3 comes from this gene; cDNA EST
yk501c3.3 comes from this gene; cDNA EST yk505d6.3 comes
from this gene; cDNA EST yk213e10.5 comes from this gene
cDNA EST yk236f7.5 comes from this gene; cDNA EST
yk241b1.5 comes from this gene; cDNA EST yk254h9.5 comes
from this gene; cDNA EST yk268g12.5 comes from this gene
cDNA EST yk278d11.5 comes from this gene; cDNA EST
yk269g3.5 comes from this gene; cDNA EST yk273e7.5 comes
from this gene; cDNA EST yk284f2.5 comes from this gene
cDNA EST yk372a2.5 comes from this gene; cDNA EST
yk335g12.5 comes from this gene; cDNA EST yk299a2.5 comes
from this gene; cDNA EST yk372f10.5 comes from this gene
cDNA EST yk391h1.5 comes from this gene; cDNA EST
yk374g7.5 comes from this gene; cDNA EST yk376a3.5 comes
from this gene; cDNA EST yk402b6.5 comes from this gene
cDNA EST yk417g2.5 comes from this gene; cDNA EST

yk425f6.5 comes from this gene; cDNA EST yk448e7.5 comes from this gene; cDNA EST yk452b3.5 comes from this gene; cDNA EST yk452d7.5 comes from this gene; cDNA EST yk454e3.5 comes from this gene; cDNA EST yk462c4.5 comes from this gene; cDNA EST yk474a2.5 comes from this gene; cDNA EST yk409h4.5 comes from this gene; cDNA EST yk437f9.5 comes from this gene; cDNA EST yk446g3.5 comes from this gene; cDNA EST yk475e1o.5 comes from this gene; cDNA EST yk483g3.5 comes from this gene; cDNA EST yk485e8.5 comes from this gene; cDNA EST yk627a7.3 comes from this gene; cDNA EST yk627a7.5 comes from this gene; cDNA EST yk630e1o.3 comes from this gene; cDNA EST yk630e1o.5 comes from this gene; cDNA EST yk633f8.3 comes from this gene; cDNA EST yk633f8.5 comes from this gene; cDNA EST yk464d9.5 comes from this gene; cDNA EST yk487d3.3 comes from this gene; cDNA EST yk509h6.3 comes from this gene; cDNA EST yk513e8.3 comes from this gene; cDNA EST yk514a6.3 comes from this gene; cDNA EST yk520c7.3 comes from this gene; cDNA EST yk524a3.3 comes from this gene; cDNA EST yk531h4.3 comes from this gene; cDNA EST yk535h6.3 comes from this gene; cDNA EST yk536b1o.3 comes from this gene; cDNA EST yk538c4.3 comes from this gene; cDNA EST yk34532.3 comes from this gene; cDNA EST yk357e8.3 comes from this gene; cDNA EST yk358g12.3 comes from this gene; cDNA EST yk573f1o.3 comes from this gene; cDNA EST yk574d5.3 comes from this gene; cDNA EST yk575e6.3 comes from this gene; cDNA EST yk605e1l.3 comes from this gene; cDNA EST yk607f8.3 comes from this gene; cDNA EST yk613b8.3 comes from this gene; cDNA EST yk614h8.3 comes from this gene; cDNA EST yk646e9.3 comes from this gene; cDNA EST yk652e4.3 comes from this gene; cDNA EST yk654e5.3 comes from this gene; cDNA EST yk663e2.3 comes from this gene; cDNA EST yk671l5.3 comes from this gene; cDNA EST yk487d3.5 comes from this gene; cDNA EST yk501c3.5 comes from this gene; cDNA EST yk505d6.5 comes from this gene; cDNA EST yk505g1o.5 comes from this gene; cDNA EST yk509h6.5 comes from this gene; cDNA EST yk513e8.5 comes from this gene; cDNA EST yk514e6.5 comes from this gene; cDNA EST yk520c7.5 comes from this gene; cDNA EST yk524a3.5 comes from this gene; cDNA EST yk531h4.5 comes from this gene; cDNA EST yk535h6.5 comes from this gene; cDNA EST yk536b1o.5 comes from this gene; cDNA EST yk539c4.5 comes from this gene

Query Match 4.8%; Score 102; DB 3; Length 39553;
Best Local Similarity 61.3%; Pred. No. 2.7e-11;
Matches 184; Conservative 0; Mismatches 110; Indels 6; Gaps 1;

QY	698	tactgtgactgctttccagtgagggacttttgcacaacactgcaattgtataattgttc	757
Db	15856	TACTGTGATTGTTTCGCCAATGAGAGAGTTCGTGCTGACTGCAATTGCAAGGATTGTCAC	15915
QY	758	aacaacttgcatacatatattgaacggttttaagaccattaaagcattgtcttgttagaaat	817
Db	15916	AATAATATAGATACGACAGCTAGCGTTCAAAAGCCATCCGTGAGTCACATGTGACGGAAAT	15975
QY	818	ccagaagcttttcagccaaaaattgggaagggccaattggg-----caatgtccaagccc	871
Db	15976	CGAAGCGTTTCAAGCCAAAAATTGCTATTGCTCGTGGAGGTATTACCGACATCGAAGCT	16035
QY	872	cagcaacacaaagggtgcgaactcgaggaggtcaggtgcctgaagaattactcgagtgc	931
Db	16036	CTTCATCAGNAGGATGCTACTGTAAAAAGAGTGGTGTCTGANAACCTATTGTGAGTGT	16095
QY	932	tatgagggcccaaatattgtgttctctatttcgaaatgcattgggtgcaaaaattatgaa	991
Db	16096	TATGAAGCAAAAGGTCCGTGTACCGATCGATGCAAGTGCAAAGATGTCAGAATACATGAA	16155

RESULT	13		
CEY67H2			
LOCUS	CEY67H2	DNA	linear
		330724 bp	
			HTG 11-NOV-1999

DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	

COMMENT

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10

```

BASE COUNT      100816 a 56710 c 54779 g 101604 t 16815 others
.ORIGIN

```

Query Match	4.8%	Score 102;	DB 2;	Length 330724;
Best Local Similarity	61.3%;	Pred. No. 4.4e-11;		
Matches 184;	Conservative 0;	Mismatches 110;	Indels 6;	Gaps 1;
Qy 698	tactgtgactgctttgcagtgagggaacttttgaacaactgcgaattgtgaataattgttgc	757		
Db 278539	TACTGTGATTGTTTCGCCAATGAGGAGCTTCTGTGCTGACTGCAATTGCAAGGATTGTCA	278598		
Qy 758	aaacaactgcatcatgatattgaacggtttaaacgcattaaagcattgctgttgataa	817		
Db 278599	AATTAATTAGAATACGACAGCTACGGTTCAAAGGCCATCCGTCAGTCATCTTGACGGAAT	278658		
Qy 818	ccagaagctttccagccaaataattgggaaggccaaattggg-----caatgtcgaagccc	871		
Db 278659	CCGAACGCTTTCGAAGCCAAAATTTGGTATTGCTGCTGGAGGTATTACCGCATCGGAACGT	278718		
Qy 872	cagcacacaagaaggtcaactcagcaggaggtcagctgcctgaagaattactcgcagatgc	931		
Db 278719	CTTCATTCAGAAAGAGATGTCACCTGTAAAAGAGGTGTTGTCGAAAACACTATTGTGAGTGT	278778		
Qy 932	tatgaggcccaaatattgtgtcttcttatttgcacaatgcattggtgcacaaatattgaa	991		
Db 278779	TATGAACCAAAAGGTTCCGTTCCGATCCGATTCGATGCAAGTGCAAAAGGATCTCAGAATAC	278838		

RESULT	14
AC097959/c	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

Caenorhabditis elegans chromosome IV clone Y67H2, *** SEQUENCING IN
PROGRESS ***, in unordered pieces.

AL022475
AL022475.12 GI:6433825
HTG: HTGS_PHASE1.
Caenorhabditis elegans.
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peleiderinae; Caenorhabditis.
1 (bases 1 to 330724)
McMurray, A.
Direct Submission
Submitted (10-NOV-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1RQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or rwnematode.wustl.edu
On Nov 15, 1999 this sequence version replaced gi:6014374.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.

AC097959 179953 bp DNA linear HTG 20-DEC-2001
Rattus norvegicus clone CH230-114G10, *** SEQUENCING IN PROGRESS
***, 62 unordered pieces.
AC097959
AC097959.3 GI:17949939
HTG; HTGS_PHASE1.
Norway rat.
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS

1 (bases 1 to 179953)
Muzny,D.M., Adams,C., Adlo-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amarantunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Blmaga,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle-McDell,D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carrall,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kowar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Monabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peary,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sison,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

TITLE
JOURNAL
AUTHORS
JOURNAL

2 (bases 1 to 179953)
Worley,K.C.

Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:16901822.

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFSK
Center clone name: CH230-114G10
----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to findPhrapList
Consensus quality: 163172 bases at least Q40
Consensus quality: 168807 bases at least Q30
Consensus quality: 174078 bases at least Q20
Estimated insert size: 160479; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently

* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 6639: contig of 6639 bp in length
6640 6739: gap of unknown length
6740 12683: contig of 5944 bp in length
12684 12783: gap of unknown length
12784 16867: contig of 4084 bp in length
16868 16967: gap of unknown length
16968 22929: contig of 5962 bp in length
22930 23029: gap of unknown length
23030 30265: contig of 7236 bp in length
30266 30365: gap of unknown length
30366 36389: contig of 6024 bp in length
36390 36489: gap of unknown length
36490 42068: contig of 5579 bp in length
42069 42168: gap of unknown length
42169 46728: contig of 4560 bp in length
46729 46828: gap of unknown length
46829 52308: contig of 5380 bp in length
52309 52308: gap of unknown length
52309 56609: contig of 4301 bp in length
56610 56709: gap of unknown length
56710 59548: contig of 2839 bp in length
59549 59648: gap of unknown length
59649 62711: contig of 3062 bp in length
62711 62810: gap of unknown length
62811 66227: contig of 3417 bp in length
66228 66327: gap of unknown length
66328 69089: contig of 2762 bp in length
69090 69189: gap of unknown length
69190 72645: contig of 3456 bp in length
72646 72745: gap of unknown length
72746 76917: contig of 4072 bp in length
76918 76917: gap of unknown length
76918 80040: contig of 3123 bp in length
80041 80140: gap of unknown length
80141 83579: contig of 3439 bp in length
83580 83679: gap of unknown length
83680 86465: contig of 2786 bp in length
86466 86565: gap of unknown length
86566 90515: contig of 3950 bp in length
90516 90615: gap of unknown length
90616 94222: contig of 3607 bp in length
94223 94322: gap of unknown length
94323 97924: contig of 3602 bp in length
97925 98024: gap of unknown length
98025 98751: contig of 1727 bp in length
98752 98851: gap of unknown length
98852 102642: contig of 2791 bp in length
102643 102742: gap of unknown length
102743 106798: contig of 4056 bp in length
106799 106898: gap of unknown length
106899 110343: contig of 3445 bp in length
110344 110443: gap of unknown length
110444 112934: contig of 2491 bp in length
112935 113034: gap of unknown length
113035 116034: contig of 3000 bp in length
116035 116134: gap of unknown length
116135 119554: contig of 3420 bp in length
119555 119654: gap of unknown length
119655 122029: contig of 2375 bp in length
122030 122129: gap of unknown length
122130 125046: contig of 2917 bp in length
125047 125146: gap of unknown length
125147 127444: contig of 2298 bp in length
127445 129251: contig of 1707 bp in length
129252 129351: gap of unknown length

* NOTE: This is a 'working draft' sequence. It currently
* consists of 90 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 9985: contig of 9985 bp in length
* 9986: gap of unknown length
* 10086: contig of 5801 bp in length
* 15887: gap of unknown length
* 15888: contig of 6816 bp in length
* 22802: gap of unknown length
* 22903: contig of 5923 bp in length
* 28826: gap of unknown length
* 28925: gap of unknown length
* 34008: contig of 5083 bp in length
* 34109: gap of unknown length
* 39104: contig of 4995 bp in length
* 39203: gap of unknown length
* 45058: contig of 5855 bp in length
* 45159: gap of unknown length
* 49553: gap of unknown length
* 49653: contig of 4068 bp in length
* 53721: gap of unknown length
* 53820: contig of 3595 bp in length
* 57416: gap of unknown length
* 57515: contig of 3612 bp in length
* 61128: gap of unknown length
* 61227: contig of 2631 bp in length
* 63858: gap of unknown length
* 63959: contig of 3592 bp in length
* 67551: gap of unknown length
* 67651: contig of 2507 bp in length
* 70157: gap of unknown length
* 70257: contig of 3624 bp in length
* 73881: gap of unknown length
* 73981: contig of 4486 bp in length
* 78467: gap of unknown length
* 78568: contig of 3188 bp in length
* 81756: gap of unknown length
* 81856: contig of 3880 bp in length
* 85736: gap of unknown length
* 85835: contig of 2610 bp in length
* 88445: gap of unknown length
* 88546: contig of 2011 bp in length
* 90557: gap of unknown length
* 90657: contig of 3086 bp in length
* 93743: gap of unknown length
* 93842: contig of 2798 bp in length
* 96641: gap of unknown length
* 96741: contig of 1991 bp in length
* 98831: gap of unknown length
* 98832: contig of 2506 bp in length
* 101338: gap of unknown length
* 101437: contig of 2073 bp in length
* 103510: gap of unknown length
* 103610: contig of 2016 bp in length
* 105627: gap of unknown length
* 105726: contig of 2873 bp in length
* 108600: gap of unknown length
* 108699: contig of 2050 bp in length
* 110749: gap of unknown length
* 110849: contig of 1048 bp in length
* 110850: gap of unknown length
* 111897: contig of 2097 bp in length
* 111998: gap of unknown length
* 114094: contig of 2002 bp in length
* 114194: gap of unknown length
* 114195: contig of 2202 bp in length
* 116197: gap of unknown length
* 116297: contig of 2728 bp in length
* 118499: gap of unknown length
* 118598: contig of 2728 bp in length
* 118599: contig of 2728 bp in length
```

```
* 121327: gap of unknown length
* 121427: contig of 1636 bp in length
* 123062: gap of unknown length
* 123162: contig of 1601 bp in length
* 124763: gap of unknown length
* 124863: contig of 2060 bp in length
* 126923: gap of unknown length
* 126924: contig of 2455 bp in length
* 127024: gap of unknown length
* 129479: gap of unknown length
* 129578: contig of 1943 bp in length
* 131521: gap of unknown length
* 131621: contig of 1728 bp in length
* 133349: gap of unknown length
* 133449: contig of 2124 bp in length
* 133450: gap of unknown length
* 135573: contig of 1573 bp in length
* 135674: gap of unknown length
* 137247: gap of unknown length
* 137346: contig of 1294 bp in length
* 138641: gap of unknown length
* 138740: contig of 1032 bp in length
* 138741: gap of unknown length
* 139772: contig of 1304 bp in length
* 139873: gap of unknown length
* 141177: contig of 1102 bp in length
* 141277: gap of unknown length
* 142379: contig of 1592 bp in length
* 144071: gap of unknown length
* 144171: contig of 2520 bp in length
* 146691: gap of unknown length
* 146791: contig of 3100 bp in length
* 149891: gap of unknown length
* 149991: contig of 1509 bp in length
* 151500: gap of unknown length
* 151600: contig of 2657 bp in length
* 154257: gap of unknown length
* 154357: contig of 2203 bp in length
* 156560: gap of unknown length
* 156659: contig of 1266 bp in length
* 157925: gap of unknown length
* 158026: contig of 1266 bp in length
* 159291: gap of unknown length
* 159391: contig of 1752 bp in length
* 161144: gap of unknown length
* 161243: contig of 1651 bp in length
* 161244: gap of unknown length
* 162894: contig of 2170 bp in length
* 162895: gap of unknown length
* 165165: contig of 2170 bp in length
* 165264: gap of unknown length
```

Query Match 3.9%; Score 83; DB 2; Length 214055;

Best Local Similarity 72.8%; Pred. No. 5.3e-07;

Matches 107; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 792 ccattagggcatgtcttggtagaaatccagaagcttccagccaaatgtgggaaggcc 851

Db 15210 CCCAAAGGCGATCCCTTGACAGAAACCCGGAAGCCTTCAAGCCAAAGATAGGAAAGGAA 15151

Qy 852 aattggcaatgtcaagcccccagcaacaagggtgaactgcaggaggtcaggctgcc 911

Db 15150 AAGAGGGGGAGTCCGATCGGGCGCACAGCAAGGCTGTAACCTGCAACCGGTCAAGGGTGC 15091

Qy 912 tgaagaattactgcagtgctatgagg 938

Db 15090 TTAATAACTACTGCGAGTGCATGAGG 15064

Search completed: July 11, 2002, 11:36:25
Job time: 12102 sec

GenCore version 4.5
Copyright (c) 1993 - 2000. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 10:54:07 ; Search time 479.19 Seconds
(without alignments)
7646.018 Million cell updates/sec.

Title: US-09-743-237-3

Perfect score: 2134

Sequence: 1 aattcggttcaggcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

1: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1980.DAT.*
2: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1981.DAT.*
3: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1982.DAT.*
4: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1983.DAT.*
5: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1984.DAT.*
6: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1985.DAT.*
7: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1986.DAT.*
8: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1987.DAT.*
9: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1988.DAT.*
10: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1989.DAT.*
11: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1990.DAT.*
12: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1991.DAT.*
13: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1992.DAT.*
14: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1993.DAT.*
15: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1994.DAT.*
16: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1995.DAT.*
17: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1996.DAT.*
18: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1997.DAT.*
19: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1998.DAT.*
20: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA1999.DAT.*
21: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA2000.DAT.*
22: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA2001A.DAT.*
23: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA2001B.DAT.*
24: /SIDSL1/gcgdata/hold-geneseg/geneseg-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2134	100.0	2134	21	AA288156 Human testis speci
2	662	31.0	1861	21	AA288157 Mouse testis speci
3	647.8	30.4	2241	21	AA288155 Mouse testis speci
4	476.8	22.3	3495	22	AAK94268 Human full-length
5	429	20.1	445	22	AA101300 Human reproductive
6	245	11.5	250	22	AAH35563 Human colon cancer
7	203.8	9.6	3533	20	AAV72867 Human lin-54 homol
8	199.8	9.4	588	22	AAK91874 Human cDNA 5'-end
9	199.8	9.4	588	22	AAK93281 Human cDNA clone r

10	147.4	6.9	1093	23	AA82145 DNA encoding novel
11	113	5.3	3164	23	ABL10437 Drosophila melanog
12	102	4.8	1503	20	AAV72865 Caenorhabditis ele
13	69.2	3.2	8372	23	ABL10436 Drosophila melanog
14	52	2.4	986	20	AAAS1735 DNA encoding a hum
15	52	2.4	1772	21	AAF21809 Human breast and o
16	50.4	2.4	368	22	AAAS6046 Human cancer agent
17	50.4	2.4	3413	21	AAA47459 Human TANGO 239 co
18	50.4	2.4	3413	21	AAA47483 Human TANGO 239 co
19	50	2.3	3234	13	AAQ30999 Notch clone hN3k f
20	48.6	2.3	2028	20	AA81861 CDNA encoding a mu
21	48.6	2.3	2028	21	AA08010 Mouse polynucleoti
22	48.2	2.3	12405	22	AAAS5330 Chemically pretrea
23	48.2	2.3	12405	24	AA61143 Human gene regulat
24	48	2.2	207	21	AAAC98751 Human colon cancer
25	48	2.2	4100	20	AAZ32019 Human METH1 relate
26	48	2.2	4100	22	AAAC90076 LI3855 CDNA clone.
27	47.6	2.2	2607	22	AAH33503 Human colon cancer
28	47.4	2.2	1909	20	AAV84356 Nuclear matrix-ass
29	47.4	2.2	1912	14	AAQ43443 AML 1 gene. Homo
30	47.4	2.2	4193	23	AA572413 DNA encoding novel
31	47.4	2.2	4287	19	AAV20475 Human AML1/MTG8 on
32	47.4	2.2	5828	20	AA84592 AML1-MTG16 fusion
33	47.4	2.2	6056	20	AA84593 AML1-MTG16 fusion
34	47	2.2	3035	24	ABK09774 Human ovarian tumo
35	46.8	2.2	1034	21	AAZ52527 Human secreted pro
36	46.8	2.2	1181	19	AAV59803 Human secreted pro
37	46.8	2.2	1212	19	AAV59686 Human secreted pro
38	46.6	2.2	10640	22	AA03729 P. falciparum telo
39	46.4	2.2	378	22	AAI90863 Human polynucleoti
40	46.4	2.2	6695	22	AAK70340 Human immune/haema
41	46.4	2.2	6695	22	AAK82282 Human immune/haema
42	46.2	2.2	1705	20	AAV08856 Gene encoding huma
43	46.2	2.2	1705	20	AAV84616 Human secreted pro
44	46.2	2.2	1705	22	ABA83399 Human secreted pro
45	46	2.2	768	22	AAI96382 Human neuroblastom

ALIGNMENTS

RESULT 1

AAZ88156

ID AA288156 standard; cDNA; 2134 BP.

XX AA288156;

AC AA288156;

XX 25-APR-2000 (first entry)

DT Human testis specific factor tesmin encoding CDNA SEQ ID NO:3.

DE Testis specific factor; tesmin; cell death; regulation; spermatocyte;

DE differentiation regulatory factor; male germ cell regulatory actor;

KW germ cell differentiation; sterility; ss.

KW Homo sapiens.

OS Key

XX Location/Qualifiers

FT CDS

FT 407..1306

FT /*tag= a

FT /product= "tesmin"

XX WO200004147-A1.

XX PD 27-JAN-2000.

XX PF 16-JUL-1999; 99WO-JP03859.

XX PR 17-JUL-1998; 98JP-0219856.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX

Qy 1861 caggcttgaccagcagctgctttagtgagtagagcacttaaaattaccagttcttctgtgt 1920
 Db 1861 caggcttgaccagcagctgctttagtgagtagagcacttaaaattaccagttcttctgtgt 1920
 Qy 1921 ttctactgaatttgaaagcctctatttcccaattaaactctccataattatgttgt 1980
 Db 1921 ttctactgaatttgaaagcctctatttcccaattaaactctccataattatgttgt 1980
 Qy 1981 aatattattttgtttaaacaatggttcacataactagcttgtgaaaccagcaggtta 2040
 Db 1981 aatattattttgtttaaacaatggttcacataactagcttgtgaaaccagcaggtta 2040
 Qy 2041 aaatgaattcttaagtgcagccttttggcttctgtttaaagcaaatgaataaaattt 2100
 Db 2041 aaatgaattcttaagtgcagccttttggcttctgtttaaagcaaatgaataaaattt 2100
 Qy 2101 ccaatgtcgaataaaataaaataaaataaaataaaataaaataaaataaaataaa 2134
 Db 2101 ccaatgtcgaataaaataaaataaaataaaataaaataaaataaaataaaataaa 2134

RESULT 2

AAZ88157 standard; cDNA; 1861 BP.

AAZ88157;

25-APR-2000 (first entry)

Mouse testis specific factor tesmin encoding cDNA SEQ ID NO:2.

Testis specific factor; tesmin; cell death; regulation; spermatocyte; differentiation regulatory factor; male germ cell regulatory actor; germ cell differentiation; sterility; ss.

Mus musculus.

Key Location/Qualifiers
 CDS 271..1158
 /*tag= a
 /product= "tesmin"

W0200004147-A1.

27-JAN-2000.

16-JUL-1999; 99WO-JP03859.

17-JUL-1998; 98JP-0219856.

(CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

WPI; 2000-147785/13.

P-PSDB; AAY68463.

New male germ cell regulatory factor tesmin expressed in spermatocytes useful for investigation of germ cell differentiation and sterility -

Example 2; Page 38-42; 63pp; Japanese.

The present sequence encodes a male germ cell regulatory factor expressed specifically in spermatocytes, designated tesmin. Tesmin can be used in the investigation of the mechanisms of germ cell differentiation and sterility.

Sequence 1861 BP; 456 A; 481 C; 498 G; 426 T; 0 other;

Query Match 31.0%; Score 662; DB 21; Length 1861;
 Best Local Similarity 74.7%; Pred. No. 1.6e-147;

	Matches	863;	Conservative	0;	Mismatches	280;	Indels	13;	Gaps	2;
Qy	183	gtgttttgcctcctggg	gcgcctggtt	cctgcgaaggagcctcc	ccaccggcgctccgcat	242				
Db	48	gtcgttacttcccggtg	ctcgcgcgcgcgtgt	gttgcgttaagcgagagtcgcgt	107					
Qy	243	gatccccgttgaaatca	aggttaagcaggtg	tactactacacaaagta	ataatccggaagaag	302				
Db	108	gatccccgttgaaatca	aaag-aagcaggtg	gtgtgtccaggcggcgcctgaagacg	166					
Qy	303	caatttgcagaaattct	ttcttctcaggaaat	cctgttgcgaatcccat	tggttccaggagac	362				
Db	167	cagctttccaggccctc	ctgtcaggaatcct	gttgcgaatcccat	catccccaggag	226				
Qy	363	tagaggatgcctcctg	ctgttcttcttaaga	aagattccaaaccat	gtgtatgtccaat	422				
Db	227	cagaggaggccctcag	ctgcctcggaaaga	agactccagcccat	gtgtgttctcagc	286				
Qy	423	tgaagggggcacacaa	atgctatgtataga	caaatcttagacaaga	gaactaaagcac	482				
Db	287	tgaaggaggcgccag	atgctctgcataga	caaatcttgcgcagg	gagctcaagcgc	346				
Qy	483	tccatttggcttcctc	agtatcaagatcaaa	ataattatctacag	tcagatgcctcaaac	542				
Db	347	tccatttggcttcctc	agtatcaagatcaaa	ataattatctacag	tcagatgcctcaaac	406				
Qy	543	caatgactgttttag	tagggagatttttgc	cagcatcaacaaaat	taaatctcattacac	602				
Db	407	caatgacaactttag	tgaggagacttctgc	cagtagcgaagta	aatctcatcacac	466				
Qy	603	aacaacttgaggagc	cttaccatcgtagt	caacgggtctgtctt	ccctcgggataca	662				
Db	467	aggttgataatggag	ctctcccatcagctg	tcaatgggtcgtctt	ccctcctggacctg	526				
Qy	663	ctctccaggaccacca	aaaaataaacttgg	ctgggtactgtgactg	tgttccagtgggg	722				
Db	527	ctctgcaagggccccc	aaaaataaacttgg	ctgggtactgtgactg	tgttccagtgggg	586				
Qy	723	acttttgcacaactg	caattgtaataatt	ttgcaacaacttgc	atcatatattgaac	782				
Db	587	acttctcaacagctg	ca-----gctg	caacaactcgcctc	gagctcgagc	634				
Qy	783	ggtttaaagccattaa	ggcgcgtcttctgt	gtagaataccagaag	cttccagccaaaaattg	842				
Db	635	gctcaaaagccataa	aggcgctgtcttga	tagaatactcgaag	cttccacaaaaaattg	694				
Qy	843	ggaaggcccaattgg	gccaattgtcaagc	ccccagcaacaaag	gggtgcgaactgaggg	902				
Db	695	ggaaggccgctctgg	agctgctaaacttc	gacacagcaagggt	gcgaactgtaagcgt	754				
Qy	903	cagctgcctgaaga	aattactcagtgct	atgagggcccaaat	tgttcttctattt	962				
Db	755	cagctgcctgaaga	aattactcagtgct	atgagggcccaaat	tgttcttctattt	814				
Qy	963	gcaaatgcatgtgtg	caaaaaattataga	aaagcccaagcaag	acacataatgagca	1022				
Db	815	gcaaatgcatgtgtg	caaaaaattataga	aaagcccaagcaag	acacataatgagca	874				
Qy	1023	tgcaaaactacatgc	agactggaggttgg	aagcagccattacc	ctgccacaacgaat	1082				
Db	875	cacccactacatgc	agctgggagcttga	gcagcattatttgc	ccagcgaagt	934				
Qy	1083	ttcaggacttccag	attcagtcagtagg	cgccttctcctc	atgcctcctggagg	1142				
Db	935	tctcaggaccctcca	aaactgagaaaaat	aggcaggtctctc	ctgtatcctctggag	994				
Qy	1143	tgttggaggccacat	gcgcctgcctgtgt	ctcagggagagag	gcccgagaaagacact	1202				
Db	995	tagtggaggccacat	gtgcctgcctgtgc	gccaggtggagaa	cagcagagcact	1054				
Qy	1203	gctccaaagtcctg	gcagagcagatgat	ctctggagggaat	tgttgaagggtcttat	1262				
Db	1055	gttcccaagcttgg	ctgagcagatgat	ctctggagggaat	tgttgaagggtcttat	1114				

PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0246907.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.

XX
DR WPI: 2001-465570/50.

Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition -

AA The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a coding sequence of the
CC invention.

Sequence 445 BP: 136 A; 105 C; 90 G; 111 T; 3 other; XX
S0

Qy	228	ccggggcggtccgcatgatccaggttgaaatcaaggtaagcaggtggtactactacaagta	287
Db	2	ccggggcggtccgcatgatccaggttgaaatcaagg-aagcaggtggtactactacaagta	60
Qy	288	ataatccgggaagcaactttgcagaaattcttgcaggaaatcctgttgcaagtcc	347
Db	61	ataatccgggaagcaactttgcagaaattcttgcaggaaatcctgttgcaagtcc	120

Qy	348	catcgtcccaaggaactagaggatgctcctcgtgctcgttctccttaagaagaagattcccaacccaa	407
Db	121	catcgtcccaaggaactagaggatgctcctcgtgctcgttctccttaagaagaagattcccaacccaa	180
Qy	408	tgctgatcgtcccaattgaagggggcacacaaatgctatgtatagacaattcctagaacaa	467
Db	181	tggtgatcgtcccaattgaagggggcacacaaatgctatgtatagacaattcctagaacaa	240
Qy	468	gagaactaaaaagcactccatttgggttcctcagatcaagatcaaaaaataattatctacagt	527
Db	241	gagaactaaaaagcactccatttgggttcctcagatcaagatcaaaaaataattatctacagt	300
Qy	528	cagatgtccctaaaccaatgaactgtcttagtagaggagatcttttgcacagcatcaacaaaaat	587
Db	301	cagatgtccctaaaccaatgaactgtcttagtagaggagatcttttgcacagcatcaacaaaaat	360
Qy	588	taaatcctattacacaaacacttgaggagccttaccatcgtagtcaacgggtctgctt	647
Db	361	taaatcctattacacaaacacttgaggagccttaccatcgtagtcaacgggtctgctt	420
Qy	648	tcacctcgggatcaactcttccagg	672
Db	421	tcacctcgggatcaactcttccagg	445
RESULT 6			
AAH35563			
ID	AAH35563 standard; cDNA; 250 BP.		
XX			
AC	AAH35563;		
XX			
DT	03-SEP-2001 (first entry)		
XX			
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:2645.		
XX			
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;		
KW	colorectal carcinoma; chromosome 11; ss.		
XX			
OS	Homo sapiens.		
XX			
PN	W0200122920-A2.		
XX			
PD	05-APR-2001.		
XX			
PF	28-SEP-2000; 2000WO-US26524.		
XX			
PR	29-SEP-1999; 99US-0157137.		
PR	03-NOV-1999; 99US-0163280.		
XX			
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;		
XX			
DR	WPI; 2001-235357/24.		
XX			
DR	P-PSDB; AAG76158.		
XX			
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,		
XX	useful for preventing, diagnosing and/or treating colorectal cancers -		
XX	Claim 1; Page 4257-4258; 9803pp; English.		
XX			
CC	AAH32943 to AAH37195 and AAG7788 represent human colon		
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where		
CC	the proteins are collectively known as colon cancer antigens. The colon		
CC	cancer antigens have cytostatic activity and can be used in gene		
CC	therapy and vaccine production. N and P may be used in the prevention,		
CC	diagnosis and treatment of diseases associated with inappropriate P		
CC	expression. For example, N and P may be used to treat disorders		
CC	associated with decreased expression by rectifying mutations or deletions		
CC	in a patient's genome that affect the activity of P by expressing		
CC	inactive proteins or to supplement the patients own production of P.		
CC	Additionally, N may be used to produce the colon cancer-associated Ps,		
CC	by inserting the nucleic acids into a host cell and culturing the cell		

XX	Sequence	3164 BP; 880 A; 865 C; 810 G; 609 T; 0 other;	
SQ			
	Query Match	5.3%; Score 113; DB 23; Length 3164;	
	Best Local Similarity	63.0%; Pred. No. 9.3e-17;	
	Matches 192; Conservative	0; Mismatches 110; Indels 3; Gaps	
Qy	698	tactgtgactgctttgcccagtgaggacttttgcaacaactgcaaatgttaataattgttgc 757	
Db	2565	tactgcgattgctttgcaaacgcgagttttgtcaggactgcacctgcaaggattgcttt 2624	
Qy	758	aacaactgcatcatgatattgaacgggtttaaaagccattaaagccatgtcttggtagaaat 817	
Db	2625	aacatctggactacgaagtggagcgggagcgtgtattccagctgctcgatcgtaac 2684	
Qy	818	ccagaagctttccagccaaaaattgggaaggccaaatggggcaattgtcgaagccccagcac 877	
Db	2685	cccagcgctttaaacccaaaattacgycgaccccaattcagg---tgatatgctgtcat 2741	
Qy	878	aacaaagggtgcaactgcagaggtgcaggctgcctggaagaattactgcgagtgctatgag 937	
Db	2742	aacaaagggtgcaactgcagaaagatcggcgtgcctcaagaactattgagtgctatgag 2801	
Qy	938	gcccaaatattgttcttctatttgcaaatgcatgttgccaaaaattatgaagaaagc 997	
Db	2802	gcaagattccctgctccagcatatgtaaatgcttggtgggtcgaacaatggaagaccgt 2861	
Qy	998	ccaga 1002	
Db	2862	ccgga 2866	
	RESULT 12		
	AAV72865		
ID	AAV72865	standard; cDNA; 1503 BP.	
AC	AAV72865;		
DT	29-MAR-1999	(first entry)	
DE	Caenorhabditis elegans synMuv gene lin-54 cDNA.		
XX	LIN-54; synthetic multivulvar; SynMuv; signal transduction;		
KW	animal model; tumour suppressor; retinoblastoma; cancer; cancer;		
KW	cell proliferation; gene therapy; ss.		
XX	Caenorhabditis elegans.		
XX	Key	Location/Qualifiers	
FT	CDS	17..1333	
FT		/*tag= a	
XX	W09854299-Al.		
XX	03-DEC-1998.		
XX	28-MAY-1998;	98WO-US11043.	
XX	28-MAY-1997;	97US-0047996.	
XX	(MASI) MASSACHUSETTS INST TECHNOLOGY.		
XX	Ceol C, Horvitz.HR, Lu X;		
XX	WPI: 1999-045362/04.		
DR	P-PSDB; AAW83392.		
XX	Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -		
PT	useful for treating diseases associated with altered levels of cell		
PT	proliferation, e.g. carcinomas		
XX	Claim 10; Fig 14; 70pp; English.		
PS			

XX	(PEKE) PE CORP NY.
XX	
XX	Venter JC, Adams M, Li PWD, Myers EW;
PI	
XX	WPI; 2001-656860/75.
XX	P-PSDB; ABB66333.
XX	
XX	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions .
PT	
XX	
PS	Claim 1; SEQ ID NO 25790; 2lpp + Sequence Listing; English. .
XX	
XX	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC	sequences (ABU01840-ABU16175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences .
XX	
XX	Sequence 8372 BP; 2351 A; 1921 C;1986 G; 2114 T; 0 other;
XX	

Qy	698	tactgtgaactgctgtttccagttcagtgagggaacttttgcacaacaactgcaattgttaataattgtgtgc	757
Db	587	tactgtgatgtttctccaaatggaggttctgtcgtgactgcaattgcgaagatgtgcac	646
Qy	758	aacaacttcgatcatgatattgaacggtttaaagccattaaaggcatgtctttggtagaat	817
Db	647	aataatatagaatcacgacagtcagcgttcaaaagccatccgtcagtcactgtgagcgaaat	706
Qy	818	ccagaagctttccaggccaaaattgggaaggcccaattggg-----caattgcgaagccc	871
Db	707	ccgaagcctttcaagccaaaattggtattgtcgtggaggtattacgcgacatcgaaagt	766
Qy	872	cagcacacaacaagggtgcgaactgcaggaggtcaggtgcctcgaagaatcacctgcagtgcc	931
Db	767	cttcacagaagaagtgcactgtaaaaagagtggtgtgtctgaaaaaacatttgtgagtg	826
Qy	932	tatgaggcccaaatattatgtgtttcttctatttgcaaatgcattggttgcaaaattatgaa	991
Db	827	tatgaagcaaaaggttccgtgtaccgcgtcatgcgaatgcgaagatggtcagaatactactaa	886

RESULT	13
ABL10436	
ID	ABL10436 standard; cDNA; 8372 BP.
XX	
AC	ABL10436;
XX	
DT	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster expressed polynucleotide SEQ ID NO 25790.
XX	
KW	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ss.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
PF	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
PR	

Query Match		3.28;	Score 69.2;	DB 23;	Length 8372;
Best Local Similarity		70.98;	Pred. No.	3.7e-06;	
Matches	92;	Conservative	0;	Mismatches 38;	Indels 0; Gaps 0;
Qy	862	tgtcaagccccacgacacaacaaagggtgcacctgcaggagggtcaggctgcctgaagaatta	921		
Db	6867	tgatatcgctcatatacaaaaagtgtcaactgc aaaagtcggctgcctaagaacta	6926		
Qy	922	ctgcagtgctatgagccccaatttatgttttcttatttgc aaatgc attggtgccaa	981		
Db	6927	tftgtagtgtatgaggcaaagattcccctgtccagcatatgtaaatcg ttgtaaagt	6986		
Qy	982	aattattгаа	991		
Db	6987	aactgataаа	6996		
RESULT	14				
AAX51735	ID AAX51735 standard; DNA; 986 BP.				
XX	AAX51735;				
XX	17-JUN-1999 (first entry)				
DE	DNA encoding а human secreted protein.				
XX	Human secreted protein; cancer; immune disorder; infection;				
KW	Inflammatory disorder; skin disorder; tumour; atherosclerosis;				
KW	restenosis; autoimmune disease; Alzheimer's disease;				
KW	peripheral neuropathy; trauma; spinal cord injury; allergy;				
KW	hematopoietic disorder; skeletal disorder; neurological disorder;				
KW	arthritis disorder; asthma; immunodeficiency disease; AIDS;				
KW	transplant rejection; ss.				
OS	Homo sapiens.				
PN	WO911293-A1.				
PD	11-MAR-1999.				
PF	03-SEP-1998; 98WO-US18360.				
PR	12-SEP-1997; 97US-0058974.				

Db 1653 catggcaaaaaaaaaaaaaaaaaaaaaa 1684

Search completed: July 11, 2002, 10:54:47
Job time: 9604 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 09:12:55 ; Search time 3455.78 Seconds
(without alignments)
8334.588 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 2134
Sequence: 1 aatcggtggtcaaggcgaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	655.4	30.7	1730	11 AK015724	Mus muscu
2	545.4	25.6	842	10 B1829802	AK015724 Mus muscu
3	513.4	24.1	518	9 A1874249	B1829802 603079876
4	500.6	23.5	507	9 AW293246	A1874249 tz63d10.x
5	498.6	23.4	505	10 B1711510	AW293246 UI-H-B12-
6	480.4	22.5	487	9 AW236823	B1711510 i497g07.x
7	448.4	21.0	453	9 A1652317	AW236823 xm49e06.x
8	436.6	20.8	473	9 AW102615	A1652317 wd60a08.x
9	428.6	20.1	435	10 B1792942	AW102615 xd68h02.x
10	417.4	19.6	426	10 B1711797	B1792942 ie49e01.x
11	413.8	19.4	429	9 AA629183	B1711797 i497g07.y
12	413.2	19.4	429	9 AA846474	AA629183 h152b01.y
13	412	19.3	474	9 A1624835	AA846474 aj56d11.s
14	412	19.3	509	9 A1968123	A1624835 ts71g06.x
15	412	19.3	642	9 A1968415	A1968123 wu13d06.x
16	388.8	18.2	414	9 AA994165	A1968415 wu02h05.x
17	382.2	17.9	398	9 AA758741	AA994165 ot61a06.s
					AA758741 ah80b05.s

C 18	371.4	17.4	772	10 B1520646	B1520646 603071608
C 19	363.2	17.0	888	10 BG773009	BG773009 602721268
C 20	348.6	16.3	355	9 A1796344	A1796344 w19h08.x
C 21	342.2	16.0	363	9 A1538405	A1538405 tp53g06.x
C 22	341.4	16.0	346	10 BE501038	BE501038 7a36a03.x
C 23	332	15.6	333	9 AL040101	AL040101 DKF2p434C
C 24	310.4	14.5	442	9 A1307797	A1307797 tb28c05.x
C 25	291.8	13.7	505	10 BG088886	BG088886 H3158G12-
C 26	291.2	13.6	336	10 B1793224	B1793224 ie49e01.y
C 27	281.6	13.2	2010	11 AK015732	AK015732 Mus muscu
C 28	275.2	12.9	280	9 A1377830	A1377830 te58d11.x
C 29	273.6	12.8	280	9 AW511334	AW511334 hd45e04.x
C 30	267.8	12.5	420	10 BG272409	BG272409 nah30a05
C 31	267.2	12.5	280	9 A1827726	A1827726 wf11d07.x
C 32	259.2	12.1	264	9 AA903758	AA903758 ok66a02.s
C 33	254.8	11.9	258	9 AA970002	AA970002 op61d08.s
C 34	251.8	11.8	266	9 AF012383	AF012383 AF012383
C 35	221.4	10.4	223	10 T19339	T19339 ci2012t Tes
C 36	214	9.7	697	9 BB611478	BB611478 BB611478
C 37	208	9.7	312	9 AW270088	AW270088 xv37g12.x
C 38	190.8	8.9	574	9 A1981460	A1981460 pat.pk005
C 39	189.4	8.9	475	10 B1706999	B1706999 fq13f03.y
C 40	185.2	8.7	627	10 BG772890	BG772890 602721090
C 41	170.2	8.0	397	10 BF937310	BF937310 fm63c07.y
C 42	161.6	7.6	638	10 BF426275	BF426275 df69g02.y
C 43	159.6	7.5	768	10 BG977029	BG977029 602845588
C 44	157.6	7.4	171	9 AL040102	AL040102 DKF2p434C
C 45	150.2	7.0	704	9 AV715308	AV715308 AV715308

ALIGNMENTS

RESULT	AK015724	AK015724	1730 bp	linear	HTC 19-JAN-2002
LOCUS	Mus musculus	Mus musculus	adult male testis cDNA	RIKEN full-length enriched library, clone:4930507A18	metalothionein-like 5, testis-specific (tesmin), full insert sequence.
DEFINITION	AK015724	AK015724	1	GI:12854168	
ACCESSION	AK015724	AK015724	1	GI:12854168	
VERSION	AK015724	AK015724	1	GI:12854168	
KEYWORDS	HTC; CAP trapper.	HTC; CAP trapper.			
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:4930507A18.			
ORGANISM	Mus musculus	Mus musculus			
REFERENCE	1 (sites)	1 (sites)			
AUTHORS	Carninci,P. and Hayashizaki,Y.	Carninci,P. and Hayashizaki,Y.			
JOURNAL	High-efficiency full-length cDNA cloning	High-efficiency full-length cDNA cloning			
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)	Meth. Enzymol. 303, 19-44 (1999)			
PUBMED	99279253	99279253			
REFERENCE	2 (sites)	2 (sites)			
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.			
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes			
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)	Genome Res. 10 (10), 1617-1630 (2000)			
MEDLINE	20499374	20499374			
PUBMED	11042159	11042159			
REFERENCE	3 (sites)	3 (sites)			
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi,N., Ishi,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi,N., Ishi,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.			
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer			
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)	Genome Res. 10 (11), 1757-1771 (2000)			


```

/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Qv35"
/tissue_type="tumor, 5 pooled (see description)"
/lab_host="DH10B"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; This library represents the normalized
version of NCI_CGAP_Qv23. Cloned unidirectionally.
Primer: Oligo dt. Average insert size 0.86 Kb. Tumor
types include: mixed Mullerian tumor, papillary serous,
clear cell, spindle cell. All are primary tumors,
metastasis positive. Constructed by Life Technologies."
BASE COUNT 169 a 106 c 103 g 140 t
ORIGIN

Query Match 24.1%; Score 513.4; DB 9; Length 518;
Best Local Similarity 99.8%; Pred. No. 1.7e-67;
Matches 514; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1594 cgagagctcaactctgtggaactctctctcagagaatgttctgtgaggtgtctgtg 1653
|||||
Db 518 CGAGAGCTCACTCTGGTCACTCTCCCTCAGAGATGTTCTCTGGAGGCTGCTCTG 459

QY 1654 catgaaccctcaatgggtttctgtgttttttcaaatatttagaataaagttctccgg 1713
|||||
Db 458 CATGAAACCCCTAATGGTCTCTCTGTGTTGTTTTCATATTTAGAAATAAGTCTCCGG 399

QY 1714 atgggtgtgtataccacttaaatctctagagaactactgaacacctaaagatttc 1773
|||||
Db 398 ATGGGCTGTGTGATACCACTTAAATCTCTAGAGAACTACTGAACACCTTAAAGATTTTC 339

QY 1774 ttagcgtatattctccagagacacgcgaactgtcagctcttcttaagggccccggg 1833
|||||
Db 338 TGTAGCTAGATATTTCCCGAGGACGCGCACTGCTCACTCTTCTTAAGGCCCGCGG 279

QY 1834 agacgagcgaatggggccctgcagggccaggtgcaccagcatgtctttaggttagagga 1893
|||||
Db 278 AGACGAGGCAATGGGCGCTCGCAGGCGCAGGCTTGACAGCATGCTCTGAGTTAGAGGA 219

QY 1894 cttaaaattaccagttctctgtgtttcttctactgaattgtgaaagctctattatcc 1953
|||||
Db 218 CTTAAATATATCCAGTTCTCTGTGTTTCTACTTGAATTTGTGAAAGCTCTATTATCC 159

QY 1954 aattaactctccataattattgttgaataattattattgtttgttaaaacatggttca 2013
|||||
Db 158 AATTAATCTCCATATATTGTTGTAATATATTATTATTGTTGTAACATGTTTCA 99

QY 2014 taactagctgtggaaccgagcggtaaaatgaattcttaagttgacgctttgtgtctg 2073
|||||
Db 98 TAAGTAGCTGTGTGAAACACGAGGTAAATGAATTTCTTAAGTTGAGGCTTTTGTGTTCTG 39

QY 2074 ttgtaagcaagatgaataaaatttccaatgc 2108
|||||
Db 38 TTGTAAAGCAAGATGAATAAAATTTTCCAATGTC 4

RESULT 4
AW293246/c 507 bp mRNA linear EST 16-JAN-2000
LOCUS
DEFINITION
UI-H-B12-ahl-a-10-0-UI.s1 NCI_CGAP_Sub4 Homo sapiens cDNA clone
IMAGE:2727067 3', mRNA sequence.
ACCESSION
AW293246
VERSION
AW293246.1 GI:6699968
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 507)
AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

```

```

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
The sequence contained an oligo-dt track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbrrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1. 507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2727067"
/clone_lib="NCI_CGAP_Sub4"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pYT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NCI_CGAP Sub4 library is a subtracted library derived from
the NCI_CGAP Sub2 library which is a subtracted library
derived from the NCI_CGAP Sub1 library, which is a
subtracted library derived from BI. BI constitutes a
mixture of 21 normalized or subtracted NCI_CGAP
libraries: NCI_CGAP_Co4, NCI_CGAP_Pr22, NCI_CGAP_Pr28,
NCI_CGAP_Co10, NCI_CGAP_Co16, NCI_CGAP_Kid5,
NCI_CGAP_Kid12, NCI_CGAP_Kid3, NCI_CGAP_Kid11,
NCI_CGAP_Lym2, NCI_CGAP_Br2, NCI_CGAP_Co8, NCI_CGAP_GLL1,
NCI_CGAP_Lu12, NCI_CGAP_Brn23, NCI_CGAP_Lu5
NCI_CGAP_Lu24, NCI_CGAP_Lu19, NCI_CGAP_GC4, NCI_CGAP_GC6,
NCI_CGAP_Brn25. These 21 libraries were pooled and a
single-stranded DNA preparation of the resulting mixture
was used as a tracer in a subtractive hybridization with
a driver whose composition is detailed below:
NCI_CGAP_Kid3 pool 1 : LLAM 3334-3337, 3682-3683,
3798-3803 (IMAGE Clones 1322376-1323911,
1456008-1456775, 1500552-1502855) NCI_CGAP_Kid5 pool 1 :
LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clones
1323912-1325831, 1471368-1472903, 1492104-1493255)
NCI_CGAP_Lu5 pool 1 : LLAM 3575-3582, 3851-3854 (IMAGE
Clones 1414920-1417991, 1520904-1522439) NCI_CGAP_GC4
pool 1 : LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
Clones 1257096-1258631, 1469084-1470983, 1475552-1476743
) NCI_CGAP_Pr22 pool 1 : LLAM 2457-2459, 2758-2759,
3062-3068 (IMAGE Clones 985608-986759, 1101192-1101959,
1217928-1220615) NCI_CGAP_Co10 pool 1 : LLAM 2644-2653,
2871-2872 (IMAGE Clones 1057416-1061255, 1144584-1145351
) Subtraction was performed as previously described
[Nonaldo, Lennon & Soares (1996): Normalization and
Subtraction: Two Approaches To Facilitate Gene Discovery.
Genome Research 6, 791-806.]
TAG_LIB=NCI_CGAP_Lu5
TAG_TISSUE=Lung
TAG_SEQ=CAAC
BASE COUNT 162 a 99 c 93 g 153 t
ORIGIN

```

```

Query Match 23.5%; Score 500.6; DB 9; Length 507;
Best Local Similarity 99.2%; Pred. No. 1.4e-65;
Matches 503; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1622 tctcagagaatgttgcctgaggtgctctgcataaaacctaatgtttctgttg 1681
|||||
Db 507 TCTCAGAGATGTGCTCTCGAGGCTGCTCTGCATGAACCCCTAATGTTCTGTTG 448

QY 1682 ttttcaaatatttagaataaagttctccggtgtgtgtgtgtgtgtgtgtgtgtgtgt 1741
|||||
Db 447 TTTTCAAAATTTAGAAATAAGTTCTCCGGATGGGCTGTTGTGTATACCACTTAAATC 388

QY 1742 tctagagaactactgaacacctaaagattttctgtgctagatatttccccagagacac 1801

```

Db	387	TTCTAGAGNACTACTGACACACCTAAAGAATTTCTGTAGCGTAGATATTTCCTCCAGAGGCAC	328
Qy	1802	gcgaactgtcagtccttctcctaaggccccccggagacgcagcgaataaggggcctcgcagggcc	1861
Db	327	GCGAACCTGTCACTTTCCTAAGGCCCCGGGAGACGCAGGCAATGGGGCCTCGCAGGCC	268
Qy	1862	aggcttcaccagcagatgtcttggatigagagcactaaaaattatccaggttctctctgtgtt	1921
Db	267	AGGCTTCGACACGAGCATGTCTTGAGTTGAGAGACTTAAATATTATCCAGTTTCTCTGTGTGT	208
Qy	1922	tctacttgaattgtggaaaagctctattatccaattaaactctcccatattattgttgta	1981
Db	207	TCTACTTTGAATTGTGGAAAAGCTCTATTATCCAAATTAACTTCTCCATAATTATTGTGTGA	148
Qy	1982	atatattattgtttgttaaaacatagtttcacataactagctgttggaaacccagcaggttaa	2041
Db	147	ATATTATTATTGTGTGTGTAACAATGGTTTCACATAACTAGCTGTGTGGAACCCAGCAGGTAA	88
Qy	2042	aatgaattctctaagttgacgcgttttggttctgttgttaagcaaaagatgaataaaaatttc	2101
Db	87	AATGAATCTTAAAGTTGACGCGTTTGGTCTCTGTGTGTAAGGCAAGCATGAATATAAAATTTTC	28
Qy	2102	caatgtcgaaaaaaataaaaaaataaaaaa	2128
Db	27	CAATGTCTTCAAAAAAATAAAAAAATAAAAAA	1

RESULT	5
BI711510/c	
LOCUS	BI711510 505 bp mRNA linear EST 19-SEP-2001
DEFINITION	Id97g07.x1 Human insulinoma Homo sapiens cdna 3', mRNA sequence.

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 505)
AUTHORS	Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearcce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wyllie,T., Martin,G., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronkoi,I., Bennett,J., Cardenas ,M., Gibbons,M., McCann,R., Cole,R., Tsagarieishvili,R., Williams,T , Jackson,Y. and Bowers,Y.
TITLE	Endocrine Pancreas Consortium
JOURNAL	Unpublished (2000)
COMMENT	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue Endocrine Pancreas Consortium Harvard University, Howard Hughes Medical Institute Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

```

FEATURES
source
      align quantity sequence stop: 463.
      Location/Qualifiers
        1..505
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone_lib="Human insulinoma"
          /tissue_type="insulinoma"
          /lab_host="DH10B (phage-resistant)"
          /note="Organ: pancreas; Vector: pBluescript SK-; Site:1:
XhoI; Site:2: EcoRI; Constructed with lambda ZAPRI system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to
pBluescript SK- by Dr. H. Inoue following the Washington

```

University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Hiroshi Inoue, MD/PhD for further
information on this library (Metabolism Division, Permutt
Laboratory, Washington University School of Medicine, Box
8127, 660 S Euclid Ave. St. Louis, MO 63110). Note: this
is a Washington University Pancreas EST project library. "

	Query Match	23.4%	Score 498.6	DB 10	Length 505
	Best Local Similarity	99.2%	Pred. No. 2.8e-65		
	Matches 501	Conservative 0	Mismatches 4	Indels 0	Gaps 0
QY	1624	tcagagaattgtctctggaggctgctcatgaaacccctaagtgtctctgtttgtt	1683		
Db	505	TCAGAGAATTGTCTCTGGAGGCTGCTCATGAAACCCCTAATGGTGTCTCTGTTGTT	446		
QY	1684	tttcaaatattttagaataaagtgtctccggatggcgtgtgtgataccacttcaaatctc	1743		
Db	445	TTTCAAAATTATTAGAAAAAAGTTCTCCGGATGGCTGTTGTGTATCCACTTAAAAATCTC	386		
QY	1744	tagaagaactactgaacacctaagatttctgtagcgtagatatctcccagagacacgc	1803		
Db	385	TAGAGAACTACTGAACACCTAAAGATTTCTGTAGCGTAGATATTCCCCAGAGGCACGC	326		
QY	1804	gaactgtcagctttctctaagcccccggagacgcagcgaatggggccctcgcagccag	1863		
Db	325	GAACTGTCACTGCTTCTCTAAGGCCCCGGGAGACGCAGGCAATGGGGCTCGCAGGCCAG	266		
QY	1864	gcttgaccagcagtctgtgagttgagggacttaaaattatccagtttctctgtgttcc	1923		
Db	265	GCTTGCAACAGCATGCTTGTAGTTAGAGGACTTAAANATTATCCAGTTCTCTGTGTTC	206		
QY	1924	tacttgaattgtggaaagctcttatatccaattaaactctccataattattgttgtaat	1983		
Db	205	TACTTGAATTTGTGGAAAGCTCTATTATCCAATTTAACTTCTCCATAATTTATTGTTGTAAT	146		
QY	1984	attattattgttgtaaaacatggttcacataactagctgtggaaaccagcaggtataaa	2043		
Db	145	ATTATTATTGTTGTAAACCATGGTTCCATAACTAGCTGTGTGGAAACCCAGCAGGTAAAA	86		
QY	2044	tgaattcttaagttgacgccttttggctctgtgttaaagcaaaagatgaataaaatttcca	2103		
Db	85	TGAATCTTAAGTTGACGCTTTTGGTCTGTGTGTAAAGCAAAAGATGAATAAAATTTC	26		
QY	2104	atgtcgaaaaaataaaaaaataaaaaa 2128			
Db	25	ATGCTCTTCAAAAAAATAAAAAAATAAAAAA 1			

RESULT	6
AW236823/c	
LOCUS	AW236823.c
DEFINITION	xm49e06.x1 NCI_CGAP GC6 Homo sapiens cDNA clone IMAGE:2687554 3', mRNA sequence.
ACCESSION	AW236823
VERSION	AW236823.1 GI:6569212
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 487)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgaps- r_email.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 457.

FEATURES

source
1. .487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2687554"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonids 1257096-1258631, 1469064-1470983, and 147592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 160 a 99 c 92 g 136 t
ORIGIN

Query Match 22.5%; Score 480.4; DB 9; Length 487;
Best Local Similarity 99.8%; Pred. No. 1.5e-62;
Matches 481; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1627 gagaatgttctctgaggctgtctgcatagaacccataatggtttctgtttttt 1886
Db 486 GAGAATGTTCTCTGGAGGCTGTCTGCATGAACCCCTAATGTTCTGTTGTTT 427
QY 1687 caaattattagaataagttctccgagtggtctgtgtatcaccacttaaatctctag 1746
Db 426 CAAATATTATGAAATPAGTTCCTCGGATGGCTGTGTGATACCACTTAAATCTCTAG 367
QY 1747 agaactactgaacacccataagattttctgtagcgttagatatttccccagagacgcgaa 1806
Db 366 AGAACTACTGAACACCTAAAGATTTCTGTAGCGTAGATATTTCCCCAGAGCGACGGA 307
QY 1807 ctgtcagttcttcttaagcccccgggagacgcagcgaatggggcctcgagccaggct 1866
Db 306 CTGTCACTCTTCTTAAGGCCCGGGAGACCGAGGCAATGGGGCCCTCGAGGCCAGGCT 247
QY 1867 tgaccagcagtccttgagtaggaacttaaaattatccagttcttctgtttctac 1926
Db 246 TGACACAGCATGCTTGAGTTAGAGACTTAAATATTCAGTTCTCTGTGTTCTAC 187
QY 1927 ttgaattgtgaaagcgtctattatccaatttaactcttccataattattgttgaattt 1986
Db 186 TTGAATTGTGAAAGCTCTATTATCCAATTAATCTCCCAATAATTATGTTGTAATATT 127
QY 1987 attatgtttgaaaaacataggttcacataactagctgttggaacacagcaggttaaatga 2046
Db 126 ATTATTGTTTGTAAAACATGGTTTACATAACTAGCTTGTGGAACCCAGCGGTAATGA 67
QY 2047 attcttaagtgcgcgttttggcttctgtttaaagaaaagatgaataaatttccaatg 2106
Db 66 ATTCTTAAGTTGACGCTTTTGGTTCTGTTGTAAGCAAGATGAATATAAATTTCCATG 7
QY 2107 tc 2108
Db 6 TC 5
RESULT 7
AI652317/c

LOCUS

DEFINITION
wb60a08.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2310038 3', mRNA sequence.
ACCESSION
AI652317
VERSION
AI652317.1 GI:4736296
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi.
1 (bases 1 to 453)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 422 Std Error: 0.00
Seq primer: -40UP from Gibco.

FEATURES

source
1. .453
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2310038"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonids 1257096-1258631, 1469064-1470983, and 147592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 151 a 88 c 85 g 129 t
ORIGIN

Query Match 21.0%; Score 448.4; DB 9; Length 453;
Best Local Similarity 99.8%; Pred. No. 9.2e-58;
Matches 449; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1659 aaaccctaagtgttctgttttttcaattatttagaataagttctcgatggg 1718
Db 453 AAAACCCCTAATGTTCTGTTGTTTTCATATTTAGAAAAATAGTTCTCCGGATGGG 394
QY 1719 ctgttgatcacacttaaaatctctagagaactactgaacacactaaagatttctgtag 1778
Db 393 CTGTTGTGATACCACCTTAAATCTTAGAGAACCTACTGAACACCTTAAGATTTCTGTAG 334
QY 1779 cgtgatatttccccagagacgcgcgaactgtcagttcttcttaagcccccgagacg 1838
Db 333 CGTAGATATTTCCCGAGGACGCGAAGCTGCAGTCTTTCCTTAAGGCCCGCGGAGACG 274
QY 1839 caggcaatggggcctcgagccaggctgtcacacagcatgtcttgagttagagacttaa 1898
Db 273 CAGGCAATGGGGCTCTCGACGCCAGGCTTCACACAGCATGCTCTGTAGTTAGAGACTTAA 214
QY 1899 aattatccagtttctctgttcttacttgatgtggaaaagctctattatccaatta 1958
Db 213 AATTATCCAGTTTCTCTGTGTTTCTACTTGAATTTGTGGAAGAGCTCTATTATCCAATTA 154

Db 353 CCTAAA-ATTTCCTGTAGCTAGATATTTCCCGAGAGCGACGCAACTGTGCAGTCTTCC 295

QY 1821 taagggcccccggagagcgcaggcaatgggctcgccaggccaggtctgcaccagcatgctc 1880

Db 294 TAAGGCCCCCGGAGACGCCAGGCANTGGGGCTCGCAGGCGAGCGCTGACCCAGCATGTC 235

QY 1881 ttgagttagaggacttaaaattatccagtttctctctgtgttcttacttgaaatttgga 1940

Db 234 TTGAGTTAGAGGACTTAAAAATATCCAGTTTCTCTCTGTGTCTTGAATTGTGGAAA 175

QY 1941 agctctattatccaaattaaacttctccataattattgttgtaataattattgtttgttaa 2000

Db 174 AGCTCTATTATCCAAATTAACCTTCCCAATAATTATGTGTGAATATTAATTATGTTGTAA 115

QY 2001 aacatggtcacataactagctgttggaaccagcaggtgaaatgaattcttaagtgcac 2060

Db 114 AACATGGTTCACATACTAGCTGTGGAAACCACGAGTAATGAATCTTAAGTTGAC 55

QY 2061 gctttggttctgttgtaaagcaaatgaataaaaaatttccaatgtcgaaaa 2114

Db 54 GCTTTTGGTCTGTGTGTAAAGCAAAAGATGAATAAAATTTCCAATGTCTTCAAA 1

RESULT 9

BI792942/c

LOCUS

DEFINITION

ie49601.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens

CDNA 3', mRNA sequence.

ACCESSION

BI792942

VERSION

BI792942.1 GI:15820667

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 435)

AUTHORS

Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T., Jackson, Y., and Bowers, Y.

TITLE

Endocrine Pancreas Consortium

JOURNAL

Unpublished (2000)

COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biochem.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brownjefas.harvard.edu)

Seq primer: -400P from Gibco

High quality sequence stop: 435.

Location/Qualifiers

1. 435

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"

/sex="Both"

/tissue_type="Islets of Langerhans"

/lab_host="DH10B"

/dev_stage="Adult"

/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Not 1; Site: 2; Sal 1; Starting library constructed using SuperScript plasmid Library kit (Life Technologies). cDNA made by oligo-dr priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from

SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 419)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: crapsb-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 403.

FEATURES
Location/Qualifiers
1..419
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2975881"
/lab_host="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CCAP-GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo.

BASE COUNT 134 a 83 c 80 g 122 t
ORIGIN

Query Match 19.4%; Score 413.8; DB 9; Length 419;
Best Local Similarity 99.5%; Pred. No. 1.4e-52;
Matches 415; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1692 tatttagaaataagttctccggatgggtgtgtgataccacttaaaatctctagagaac 1751
|||||
Db 419 TATTTAGAAATAAGTTCTCCGATGGGCTGTGTGATACCACCTTAAATCTCTAGAGAAC 360
|||||
Qy 1752 tactgaacacctaagaattttctgttagcgtagatatctcccgagacacgcgaactgtc 1811
|||||
Db 359 TACTGAACACCTAAAGATTTTCTGTAGGCTAGATATTTCCCGAGGACGCGAACTGTC 300
|||||
Qy 1812 agtcttcttaagcccccgagacgcagcaaatggggccctcgcagccaggttcgac 1871
|||||
Db 299 AGTCTTCTTAAGGCCCGGAGACGAGCAATGGGGCCCTCGCAGGCCAGGCTTGAC 240
|||||
Qy 1872 cagcatgtctgagtagaggaacttaaaattatccagttctctgtgttcttacttgaa 1931
|||||
Db 239 CAGCATGCTTGAGTTAGAGGACTTAAATATATCCAGTTCTTCGTCTTACTTGA 180
|||||
Qy 1932 ttgtggaagactctattatccaataactctctccataattattgttgaatattat 1991
|||||
Db 179 TTGTGGAAGAGCTCTATTTATCCAAATTAATTTCTCCATATTTATTTATTTAT 120
|||||
Qy 1992 tttgttgaacatggttcacataactagctgtgtggaaccagcaggttaaatgaattct 2051
|||||
Db 119 TGTTTGTAAACATGTTTACATAACTAGCTGTGTGGAACCCAGCAGGTAATGAATCT 60
|||||
Qy 2052 taagttagcgtctttgtgtctgtgtgaaagcaaatgaataaaatttccaatgtc 2108
|||||
Db 59 TAAGTTGAGCTTTTGGTCTGTGTTAAAGCAACAATGAATAAATAATTTCCAATGTC 3

RESULT 12
AA846474/c

LOCUS AA846474
DEFINITION aj56d11.sl Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1394325
3', mRNA sequence.
ACCESSION AA846474
VERSION AA846474.1 GI:2933614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 429)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: crapsb-r@mail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo,
Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 432 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 413.

FEATURES
Location/Qualifiers
1..429
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1394325"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 140 a 83 c 79 g 127 t
ORIGIN

Query Match 19.4%; Score 413.2; DB 9; Length 429;
Best Local Similarity 99.1%; Pred. No. 1.7e-52;
Matches 426; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 1682 ttttcaaatatttagaataaagttctccggatgggtgtgtgataccacttaaaatc 1741
|||||
Db 429 TTTTATAATTTAGAAATAAGTTCTCCGATGGGCTGTGTGATACCACCTTAAATC 370
|||||
Qy 1742 tctagaactactgaacacacctaagaattttctgttagcgtagatatctcccgagagac 1801
|||||
Db 369 TCTAGAACTACTGAACACCTTAAAGATTTCTGTAGGCTAGATATTTCCCGAGAGGCAC 310
|||||
Qy 1802 gcgaactgtcagttcttctaagccccggagcgcagcaatggggcctcgcagggcc 1861
|||||
Db 309 CGGAAGTCTCAGTCTTTCTTAGGCCCGCGGAGACGAGGCAAT- GGGCTTCGAGGCC 251
|||||
Qy 1862 aggtctgaccagcagcagttctgttagtagaggacttaaaattatccagttctctgtgt 1921
|||||
Db 250 AGGCTTGACACGACGATGCTTGTAGTTAGAGGACTTAAAAATTATCCAGTTCTCTGTGT 191
|||||
Qy 1922 tctactgaattgtgaaagactctattatccaatttaactctccataattattgttga 1981
|||||
Db 190 TCTACTTGAATGTGAAAGCTCTATTATCAATTAATTAATTAATTAATTAATTAATTA 131

Qy	1982	atattattatgtttgtaaacaatggtttacataactagcttggtagaaacacgagtaa	2041
Db	130	ATATTATTATTTGTTGTAAACATGTTTCATAACTAGCTTGTGGAACACGACAGTAA	71
Qy	2042	aatgaattcttaagttagcgttttggcttctgtgttaaacgaagaatgaataaaatttc	2101
Db	70	ANYGAATCTTAGTTGACCGCTTTTGGTCTCTGTTGTTAAACGAAGAATGAATAAAATTC	11
Qy	2102	caatgtcgaa	2111
Db	10	CAATGTCAAA	1
RESULT	13		
AI624835/c			
LOCUS			
DEFINITION		AI624835	474 bp mRNA linear EST 14-DEC-1999
ACCESSION		AI624835	ts71906.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2236762 3',
VERSION		AI624835.1	mRNA sequence.
KEYWORDS		EST.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		1 (bases 1 to 474)	
JOURNAL		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
COMMENT		National Cancer Institute, Cancer Genome Anatomy Project (CGAP),	
		Tumor Gene Index	
		Unpublished (1997)	
		Contact: Robert Strausberg, Ph.D.	
		Email: cgaps-r@mail.nih.gov	
		Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael	
		R. Emmert-Buck, M.D., Ph.D.	
		cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima	
		Bonaldo, Ph.D.	
		cDNA Library Arrayed by: Greg Lennon, Ph.D.	
		DNA Sequencing by: Washington University Genome Sequencing Center	
		Clone distribution: NCI-CGAP clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LLNL at:	
		www-bio.llnl.gov/bbrp/image/image.html	
		Insert length: 878 Std Error: 0.00	
		Seq primer: -400P from Gibco	
		High quality sequence stop: 455	
FEATURES		POLYA=No.	
source		Location/Qualifiers	
		1..474	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:2236762"	
		/clone_lib="NCI_CGAP_GC6"	
		/tissue_type="pooled germ cell tumors"	
		/lab_host="DH10B"	
		/note="Vector: pT73D-Pac (Pharmacia) with a modified	
		polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA	
		from the normalized library NCI_CGAP_GC4 was prepared, and	
		ss circles were made in vitro. Following HAP purification,	
		this DNA was used as tracer in a subtractive hybridization	
		reaction. The driver was PCR-amplified cDNAs from a pool	
		of 5,000 clones made from the same library (cloneIDs	
		1257096-1258631, 1489064-1470983, and 1475592-1476743).	
		Subtraction by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT		141 a	102 c
ORIGIN		92 g	139 t
Query Match		19.38;	Score 412; DB 9; Length 474;
Best Local Similarity		98.88;	Pred. No. 2.5e-52;
Matches 415; Conservative		0; Mismatches	5; Indels 0; Gaps 0;
Qy	1689	aattatttagaataaagttctccggaatggcgttggtagataccacttaaatctctagag	1748
Db	424	TTTTTGAAGAATAAAGTCTCCGATGGCTGTTGTGATACCACTTAAATCTCTAGAG	365

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 10:45:37 ; Search time 95.08 seconds
(without alignments)
5513.062 Million cell updates/sec

Title: US-09-743-237-3

Perfect score: 2134

Sequence: 1 aattcgggtcaggcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_NA:*
- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
 - 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
 - 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
 - 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
 - 5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
 - 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48.6	2.3	2028	2	US-09-211-930-12
2	48.6	2.3	2028	3	US-09-340-993-12
3	48.6	2.3	2028	4	US-09-468-442-12
4	48.4	2.3	3234	1	US-08-264-534-31
5	48.4	2.3	3234	1	US-08-083-590A-10
6	48.4	2.3	3234	1	US-08-485-500-31
7	48.4	2.3	3234	2	US-08-346-128-31
8	48.4	2.3	3234	3	US-08-532-384-10
9	48.4	2.3	3234	3	US-08-893-828-31
10	47.4	2.2	1909	3	US-09-100-193-6
11	47.4	2.2	4287	1	US-08-244-189-1
12	47.4	2.2	4287	1	US-08-306-691B-53
13	43.6	2.0	2435	4	US-09-306-593-1
14	42.8	2.0	2497	2	US-09-014-969-14
15	42.8	2.0	2993	2	US-08-415-593-42
16	42.6	2.0	6124	4	US-08-213-419B-3
17	42.4	2.0	6243	2	US-09-056-075-1
18	41.8	2.0	5923	4	US-09-064-922-3
19	41.6	1.9	5467	1	US-07-745-206A-12
20	41.6	1.9	5467	2	US-08-311-363-12
21	41.6	1.9	5852	1	US-07-867-106-2
22	41.4	1.9	739	4	US-08-791-115B-15
23	41.4	1.9	1046	1	US-08-361-467B-4
24	41.4	1.9	1046	1	US-08-484-332C-4
25	41.4	1.9	1117	4	US-09-247-373B-33
26	41.4	1.9	1990	1	US-08-256-130A-1
27	41.4	1.9	3871	2	US-08-599-455B-3

28	41.4	1.9	3871	4	US-09-069-781B-3	Sequence 3, Appl
29	41.2	1.9	1395	2	US-08-553-367A-1	Sequence 1, Appl
30	41.2	1.9	1395	4	US-09-295-306-1	Sequence 1, Appl
31	41	1.9	1493	1	US-08-340-820-24	Sequence 24, Appl
32	41	1.9	1493	1	US-08-593-535-24	Sequence 24, Appl
33	40.8	1.9	573	4	US-08-906-156A-67	Sequence 67, Appl
34	40.6	1.9	1550	2	US-08-609-443B-17	Sequence 17, Appl
35	40.6	1.9	1776	3	US-08-655-352-10	Sequence 10, Appl
36	40.4	1.9	590	4	US-08-906-156A-19	Sequence 19, Appl
37	40.4	1.9	972	5	PCT-US95-04801-2	Sequence 2, Appl
38	40.4	1.9	1101	3	US-08-987-904A-1	Sequence 1, Appl
39	40.2	1.9	1804	2	US-08-504-459-5	Sequence 5, Appl
40	40.2	1.9	6232	4	US-08-456-200B-11	Sequence 11, Appl
41	40	1.9	2065	4	US-09-370-473-5	Sequence 5, Appl
42	40	1.9	6854	4	US-09-194-905-7	Sequence 7, Appl
43	40	1.9	19124	2	US-08-487-826B-13	Sequence 13, Appl
44	39.6	1.9	3200	1	US-08-444-405-1	Sequence 1, Appl
45	39.6	1.9	3200	1	US-08-384-850-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-211-930-12
; Sequence 12, Application US/09211930
; Patent No. 5962265
; GENERAL INFORMATION:
; APPLICANT: Tyrell E. No. 5962265rls
; APPLICANT: William Craig Moore
; TITLE OF INVENTION: HUMAN SIGNAL TRANSDUCTION
; TITLE OF INVENTION: SERINE/THREONINE KINASE
; FILE REFERENCE: PHM.70296
; CURRENT APPLICATION NUMBER: US/09/211.930
; EARLIER FILING DATE: 1998-12-15
; EARLIER FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 2028
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-211-930-12

Query Match 2.3%; Score 48.6; DB 2; Length 2028;
Best Local Similarity 51.3%; Pred. No. 0.007;
Matches 138; Conservative 0; Mismatches 129; Indels 2; Gaps 1;

Qy 1868 gcacagcagtcgttgaggttagaggaactaaattatccagttttctctgtgttctact 1927
Db 1738 gtaccattattcttgcattgatgaatgactgtttgttggaactcttttagtaac 1797
Qy 1928 tgaattgtggaagactcta--ttatccaattaaactctccataattattgtgtaaat 1985
Db 1798 tgtatactcaataacatacaggtctttccaagtcattcctaactataaattgttgaat 1857
Qy 1986 tattattgttgtaaacatggttcacataactagcttctgtggaacacagcaggtaaatg 2045
Db 1858 catcaagcttcaaaaagcattctttcccccacacagatattctctaaatgactatt 1917
Qy 2046 aattcttaagtgcgcgttttgggttctgttgaagcaaaagatgaataaatttccaat 2105
Db 1918 tgaataggttggaagtaagtaatactcttcttaaacacagtggttttttaagaagctccgg 1977
Qy 2106 gtcaaaaaaaaaaaaaaaaaaaaaaaaaa 2134
Db 1978 aaaaaaaaaaaaaaaaaaaaaaaaaa 2006

RESULT 2

Query Match	2.3%;	Score 48.6;	DB 4;	Length 2028;
Best Local Similarity	51.3%;	Pred. No. 0.007;		
Matches 138;	Conservative 0;	Mismatches 129;	Indels 2;	Gaps 1;
QY 1868	gcaccagcagtctgttgagttagagagacttaaaattccacgtttctctctgtgtttctact	1927		
DB 1738	gtaccattattcttattgtttcatgtaagaatactgttggtttggaaatcttttagttaac	1797		
QY 1928	tgaatttgaaaaagctcta--ttatccaattactctctccataattattttgttaatat	1985		
DB 1798	tgtatactataacacacaggtctttcaagtcatactaaatgatttgaatgttgaat	1857		
QY 1986	tattattgtttgtaaaacatggttcacataactagcttgtggaaaccagcaggtaaaatg	2045		
DB 1858	catcaagctctcaaaagcatctcttttccccacacagtgatttctctaaaaatgactatt	1917		
QY 2046	aattcttcaagtgtgacgcttttggttctgttgtgaagcaagatgaataaaaaatttccaat	2105		
DB 1918	tgtaatgaggtggaagtaagtaaaccttctttaaaacaagtgtttttaagaagctcccg	1977		
QY 2106	gtcgaaaaaataaaaaaataaaaaa 2134			
DB 1978	aaagaaaaaataaaaaaataaaaaa 2006			

RESULT 4
US-08-264-534-31
; Sequence 31, Application US/08264534
; Patent No. 5648464
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon et al.
; TITLE OF INVENTION: Human No. 5648464ch And Delta, Binding Domains
; TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264.534

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/264,534
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/695,189
 FILING DATE: 03-MAY-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Misrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-004
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 790-9090
 TELEFAX: 212 8698864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 31:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3234 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: unknown
 MOLECULE TYPE: CDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..3234
 US-08-264-534-31

```

Query Match          2.38; Score 48.4; DB 1; Length 3234;
Best Local Similarity 55.3%; Pred. No. 0.01;
Matches 94; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy 18 aagctcgcggggcgacagcgagcggggagctcctcggggagtagtaccgccggatccca 77
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 289 AACAGCGGGAGTGGGAGTGGGAGGGCTGGACTGTGCGGAGCATGTACCGAGAGGCTG 348
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 78 gagctcagcgcgctggaggacgtcgctcgtcgagggcccgagcgcccgcttgcgaac 137
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 349 GCGGCGGCGACCGCTGGTGGTGGTGTGATGCGCGCGGAGCAGCTGCGCAACAGCTCC 408
   || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 138 gtgcacttctgtctcgtctactcccgcgacgcagcgcccggggtttt 187
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 409 TTCCACTTCCTGCGGGAGCTCAGCGCGGCTGTCACACACCAACGTGGTCTT 458
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
US-08-083-590A-10
; Sequence 10: Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennile & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3234
US-08-083-590A-10

```

```

Qy 78 gagctcagcgcggtgaggacgctgcgctctcagggcccccagccccgcgcctcgtcgaac 137
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 349 GCGCCGCGCAGCTGGTGGTGTTGCTGATGCCCGCGAGCAGCTGCACAACAGCTCC 408
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 138 gtgcacttcctcctcgtcgtctaccgcgcgccaccgcagcccgcggggttt 187
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 409 TTCCACTTCCTCGGGAGCTCACCGCGTGTCTGCACACCACCACTGGTCTT 458
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT      6
US-08-465-500-31
; Sequence 31, Application US/08465500
; Patent No. 5789195
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Muskavitch, Marc A.T.
; APPLICANT: Fehon, Richard G.
; APPLICANT: Rebay, Ilaria
; APPLICANT: Blaumuellier, Cristine M.
; APPLICANT: Shepard, Scott B.
; TITLE OF INVENTION: HUMAN NOTCH AND DELTA, BINDING DOMAINS
; TITLE OF INVENTION: IN TOPORYTHMIC PROTEINS,AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,500
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-034
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-465-500-31

```

Query Match	2.3%	Score 48.4;	DB 1;	Length 3234;
Best Local Similarity	55.3%;	Pred. No.	0.01;	
Matches	94;	Conservative	0;	Mismatches 76; Indels 0; Gaps 0;
Qy	18	aagctcgggggtgcacagcagcgagggtctcttcggggagttaaccgccgatccca	77	
Dd	289	AACAGCGGGAGTGCAGGTGGGACGGGTGCAGCTGCGGGAGCATGTACCCGAGAGGCTG	348	
Qy	78	gagctcagcgcgctggaggaaegtgcgcctcctgcagccccgcgcgcgcgccttgaac	137	
Dd	349	GCGGCCGACACTGGTGTTGGTGGTGATGCCGCCGAGCAGCTGCGCAACAGCTCC	408	
Qy	138	gtgcacttcctctcctcgctgtcacccgcgcacccgcagcccgcgggttgtt	187	
Dd	409	TTCACACTTCCTGCGGGGAGCTCAGCGCGGTGCTGCACACCACCGTGGTCTT	458	

;; TITLE OF INVENTION: IN TOPOTHYMIC PROTEINS AND METHODS BASED THEREON
;; NUMBER OF SEQUENCES: 34
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: PENNIE & EDMONDS
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10036-2711
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/893,828
;; FILING DATE: 11-JUL-1997
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Mistrock, S. Leslie
;; REGISTRATION NUMBER: 18,872
;; REFERENCE/DOCKET NUMBER: 7326-050
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-8864/9741
;; INFORMATION FOR SEQ ID NO: 31:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3234 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: CDNA
US-08-893-828-31

Query Match 2.3%; Score 48.4; DB 3; Length 3234;
Best Local Similarity 55.3%; Pred. No. 0.01;
Matches 94; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
Qy 18 aagctcgggggggacagcagcgggggagctcctcgggggagtagtaccgggagatccca 77
Db 289 AACAGCGGGAGTGGAGTGGGGGCTGGACTGTGGGGAGCATGTATACCCGAGAGGCTG 348
Qy 78 gagctcagcgcgtggagagcgtcgtcctcagcggcccgagccgcgcctgcaac 137
Db 349 GCGCGCGCACGCTGGT 408
Qy 138 gtgcaactcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 187
Db 409 TTCACACTTCTCGGGAGCTCAGCGCGCTGCTGCACACACCAACGTGGTCTT 458

RESULT 10
US-09-100-193-6
; Sequence 6, Application US/09100193
; Patent No. 6153729
; GENERAL INFORMATION:
; APPLICANT: Gary S. Stein et al.
; TITLE OF INVENTION: NUCLEAR MATRIX TARGETING PEPTIDES AND USES THEREFORE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/09/100,193
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/050,104
;; FILING DATE: 20-JUNE-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jane E. Remillard
;; REGISTRATION NUMBER: 38,872
;; REFERENCE/DOCKET NUMBER: UMM-024
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 227-7400
;; TELEFAX: (617) 742-4214
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1909 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 766..1515
US-09-100-193-6

Query Match 2.2%; Score 47.4; DB 3; Length 1909;
Best Local Similarity 55.0%; Pred. No. 0.014;
Matches 93; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
Qy 8 ggtcaaggcgaagctcgcgggggacagcagcgggggagctcctcgggggagtagtacc 67
Db 837 GATGAGCGAGGCGTGTGGCGTGGGGCGCCCGGAGCGCGCGCTGCCCTGGCGGCAAGCT 896
Qy 68 cgggataccagagctcagcgcgtggagagcgtcgtcctcagcggcccgagccgcgcgc 127
Db 897 GAGGAGCGGCGACGCGAGCATGGTGGAGGTGCTGGCCGACACCGCGGCGAGCTGGTGG 956
Qy 128 cgcctgaacgtgcacttcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 176
Db 957 CACCGACAGCCCCAATTCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005

RESULT 11
US-08-244-189-1
; Sequence 1, Application US/08244189
; Patent No. 5580727
; GENERAL INFORMATION:
; APPLICANT: Ohki, Misao
; APPLICANT: Kikuchi, Kimiko
; APPLICANT: Miyoshi, Hiroyuki
; APPLICANT: Kozu, Tomoko
; TITLE OF INVENTION: No. 5580727el DNAs, Polypeptides Encoded Thereby
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/244,189
; FILING DATE: 15-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.

REGISTRATION NUMBER: 28.977
REFERENCE/DOCKET NUMBER: 760-183P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4287 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 1579..3834
US-08-244-189-1

```
Query Match      2.2%; Score 47.4; DB 1; Length 4287;
Best Local Similarity 55.0%; Pred. No. 0.021;
Matches 93; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
```

Qy	8	ggctaaagcgaagctcgggggggcgacgacgacgagagctcctcggggagtaacc	67
Db	1650	GATGACGAGGCTTCCCTGGGGCCCGGACGCCGGCGCTGCCCTGGCCGCCAGACT	1709
Qy	68	cgggataccagagctcagcgctcgggagcgtcgctcctcagcccccgacgcgc	127
Db	1710	GAGGACGCGCACCGAGCATGTGGAGGTGCTGGCCACCACCGGGCGAGCTGGTGG	1769
Qy	128	cgcctgaacgtgcattcctctcgtcgtaccccgcaaccgacg	176
Db	1770	CACCGACAGCCCCAATCTCTGCTCGGTGCTGCTACGACACTGGCGC	1818

RESULT 12
US-08-306--691B-53
US-08-306--691B-53
Sequence 53, Application US/08306691B
Patent No. 5734039
GENERAL INFORMATION:
APPLICANT: Calabretta, Bruno
APPLICANT: Skorski, Tomasz
TITLE OF INVENTION: ANTISENSE
TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seidel, Gonda, Lavorigna & Monaco, P.C.
STREET: Two Penn Center, Suite 1800
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720 kb
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/306,691B
FILING DATE: September 15, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-8

TELECOMMUNICATION INFORMATION
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
TELEX: No. 573403DNE
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 4287 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-306-691B-53

Query Match 2.2%; Score 47.4; DB 1; Length 4287;
Best Local Similarity 55.0%; Pred. No. 0.021;
Matches 93; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

Qy	8	ggtcaagcgaagctcggggggcgacagcggcggggaagctcctcgggagatccc	67
Db	1650	GATGAGCGAGCGCTTCCCTGGCGCCCGGAGCGCGGCGCTGCCCTGGCGCGCAGCT	1709
Qy	68	cgggatccagagctcagcgctggaggactcgctctctcgaagcccgacgacgcgc	127
Db	1710	GAGGAGCGCGACCGCAGCATGTTGAGTGTGTGCCACACCACCGGCGAGCTGGTGCG	1769
Qy	128	cgctgcgaactgacttctctcctcgctgtaccccgacacgcgcacgc	176
Db	1770	CACGACGAGCCCCAACTTCTGTGTCCGTGCTGTGCTACGACTGGCGC	1818

RESULT 13

```

US-09-306-593-1/C
; Sequence 1, Application US/09306593
; Patent No. 6184018
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Ljungdahl, Lars G.
; APPLICANT: Chen, HuiZhong
; APPLICANT: Ximenes, Eduardo A.
; TITLE OF INVENTION: Beta-glucosidase Coding Sequences and Protein from
; TITLE OF INVENTION: Orpinomyces PC-2
; FILE REFERENCE: 31-98us
; CURRENT APPLICATION NUMBER: US/09/306,593
; CURRENT FILING DATE: 1999-05-06
; EARLIER APPLICATION NUMBER: US 60/084,494
; EARLIER FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2435
; TYPE: DNA
; ORGANISM: Orpinomyces sp. PC-2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (39)..(2009)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (87)..(2009)
US-09-306-593-1

```

Query Match 2.0%; Score 43.6; DB 4; Length 2435;
Best Local Similarity 50.5%; Pred. No. 0.15;
Matches 106; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

[illegible]

QY 2015 aactagcttgggaaccagcggttaaatgaattcttaagttagcgcttttggttctgt 2074
Db 2107 CACTTCTTATAGCAAAAGAGTAAATAATCCATTATTAATATATTTATTATAA 2048
QY 2075 tctaaagcaagatgaataaaatttccaa 2104
Db 2047 TTGTAATAAATATTATTATAAATTTTAA 2018

RESULT 14

US-09-014-969-14

; Sequence 14, Application US/09014969

; Patent No. 5965397

; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth

; APPLICANT: McCoy, John M.

; APPLICANT: Lavallie, Edward R.

; APPLICANT: Racie, Lisa A.

; APPLICANT: Merberg, David

; APPLICANT: Treacy, Maurice

; APPLICANT: Spaulding, Vikki

; APPLICANT: Agostino, Michael J.

; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genetics Institute, Inc.

; STREET: 87 CambridgePark Drive

; CITY: Cambridge

; STATE: MA

; COUNTRY: U.S.A.

; ZIP: 02140

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/014,969

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Sprunger, Suzanne A.

; REGISTRATION NUMBER: 41,323

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617) 498-8284

; TELEFAX: (617) 876-5851

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2447 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-09-014-969-14

Query Match 2.0%; Score 42.8; DB 2; Length 2447;

Best Local Similarity 58.3%; Pred. NO. 0.24;

Matches 91; Conservative 1; Mismatches 63; Indels 1; Gaps 1;

QY 1980 taatattatttggtaaacatggttcacataactagct-tgtggaaccagcagg 2038
Db 2152 TAATGGCTATAGATATGATATATGTTTACATATCTGGATCTGTGTTGTTTGG 2211
QY 2039 taaatgaattcttaagttgacgcttttggttctgttgaagcaagaatgaataaaat 2098
Db 2212 TACTTTAAATGTGACAAATAAACCTTTGGGAGAAAAAATAAAAAAATAAAAA 2271

QY 2099 ttccaatgtcgaataaaataaaataaaataaaataaa 2134

Db 2272 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2307

RESULT 15

US-08-415-593-42

; Sequence 42, Application US/08415593

; Patent No. 5912140

; Patent No. 5912140 5776726

; GENERAL INFORMATION:

; APPLICANT: Whoriskey, Susan K.

; APPLICANT: Quinn, Cheryl L.

; APPLICANT: Tao, Nijun

; APPLICANT: Politis-Virk, Karen I.

; APPLICANT: Schimmel, Paul R.

; TITLE OF INVENTION: Recombinant Pneumocystis Aminoacyl tRNA

; TITLE OF INVENTION: Synthetase Genes, Tester Strains and Assays

; NUMBER OF SEQUENCES: 46

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

; STREET: Two Militia Drive

; CITY: Lexington

; STATE: Massachusetts

; COUNTRY: U.S.

; ZIP: 02173

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/415,593

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Brook, David E.

; REGISTRATION NUMBER: 22,592

; REFERENCE/DOCKET NUMBER: CPI94-09

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617-861-6240

; TELEFAX: 617-862-9540

; INFORMATION FOR SEQ ID NO: 42:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2993 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..2934

US-08-415-593-42

Query Match 2.0%; Score 42.8; DB 2; Length 2993;

Best Local Similarity 51.6%; Pred. No. 0.26;

Matches 98; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1945 ctattccaatttaacttctccataattattgttggaattattatttcttgaataaaca 2004
Db 2803 CTTTAAACCAAAATCTTCGTAGGCTTTAGAAAAAATCAATCAATATATTAGTGAACA 2862
QY 2005 tgggttcacataactagcttggtaaacaccagcaggtaaaatgaattcttgaattgcgctt 2064
Db 2863 GATCCTAAACAGATTGTAAGAAAAACAGGATGTTCAAGGCGCAACATTTTATGTATCT 2922
QY 2065 tgggttcgttgaagcaagaatgaataaaatttccaaatgcgaataaaataaa 2124
Db 2923 TTATTACGATTATAAATATTTTAAATAAATATTTTAAATAAATCAATTAATATTTCAAAAAA 2982
QY 2125 aaaaaaaa 2134
Db 2983 AAAAAAAAAA 2992

us-09-743-237-3.rni

Fri Jul 12 08:35:25 2002

Search completed: July 11, 2002, 10:46:24
Job time: 9101 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 11, 2002, 12:28:35 ; Search time 5652.01 Seconds
(without alignments)
8168.576 Million cell updates/sec

Title: US-09-743-237-3
Perfect score: 2134
Sequence: 1 aattcggggtcaaggcggaag.....aaaaaaaaaaaaaaaaaaaaa 2134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 21979536 seqs, 1081749327 residues 43959072
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

- 1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
- 2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
- 3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
- 4: /cgn2_6/ptodata/2/pna/US080_COMB.seq.*
- 5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
- 6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
- 7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
- 8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
- 9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
- 10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
- 11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
- 12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
- 13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
- 14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
- 15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
- 16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
- 17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
- 18: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
- 19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
- 20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
- 21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
- 22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
- 23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
- 24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
- 25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
- 26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
- 27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
- 28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
- 29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
- 30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
- 31: /cgn2_6/ptodata/2/pna/US098A_COMB.seq.*
- 32: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
- 33: /cgn2_6/ptodata/2/pna/US098C_COMB.seq.*
- 34: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.*
- 35: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
- 36: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.*
- 37: /cgn2_6/ptodata/2/pna/US100_COMB.seq.*
- 38: /cgn2_6/ptodata/2/pna/US101_COMB.seq.*
- 39: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
- 40: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
- 41: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
- 42: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
- 43: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*

44:	/cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
45:	/cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
46:	/cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
47:	/cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
48:	/cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
49:	/cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
50:	/cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
51:	/cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
52:	/cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
53:	/cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
54:	/cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
55:	/cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
56:	/cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
57:	/cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
58:	/cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
59:	/cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
60:	/cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
61:	/cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
62:	/cgn2_6/ptodata/2/pna/US6023_COMB.seq.*
63:	/cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
64:	/cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
65:	/cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
66:	/cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
67:	/cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
68:	/cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
69:	/cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
70:	/cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
71:	/cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
72:	/cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
73:	/cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
74:	/cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
75:	/cgn2_6/ptodata/2/pna/US6036_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2134	100.0	2134	29	US-09-743-237-3
2	2101	98.5	2638	25	US-09-643-893-287
3	1631.6	76.5	1770	18	US-09-471-275-5610
c	662	31.0	1861	29	US-09-743-237-2
5	647.8	30.4	2241	29	US-09-743-237-1
6	633.8	29.7	2201	75	US-60-360-207-983
7	476.8	22.3	3495	23	US-09-611-526-2894
8	429	20.1	445	1	PCT-US01-01329-425
9	429	20.1	445	1	PCT-US01-01339-1301
10	429	20.1	445	30	US-09-764-891-1301
11	429	20.1	445	37	US-10-080-090-425
12	404	18.9	465	17	US-09-362-510-45961
13	404	18.9	465	17	US-09-362-510A-45961
14	404	18.9	465	34	US-09-904-013-45961
15	401.4	18.8	431	13	US-08-971-197-11928
16	401.4	18.8	431	13	US-08-971-197-11928
c	394.4	18.5	397	16	US-09-289-768-13239
18	394.4	18.5	397	35	US-09-939-397-13239
c	308	14.4	441	18	US-09-489-036-32385
19	308	14.4	441	35	US-09-943-143-32385
20	308	14.4	441	35	US-09-943-143-32385
21	270.8	12.7	302	14	US-09-013-812-1497
22	270.8	12.7	302	21	US-09-540-764-42635
23	270.8	12.7	302	42	US-60-036-570-1497
24	249	11.7	249	20	US-09-539-806-21625
25	245	11.5	250	1	PCT-US00-26524B-2645
26	230	10.8	444	18	US-09-404-549-314
27	230	10.8	444	18	US-09-404-549A-314
28	221.6	10.4	4276	75	US-60-360-207-19806
29	203.8	9.6	3535	14	US-09-087-136-16
30	203.8	9.6	3535	14	US-09-087-136-16
31	203.8	9.6	3535	16	US-09-220-091-16

Sequence 3, Appli
Sequence 287, App
Sequence 5610, Ap
Sequence 2, Appli
Sequence 1, Appli
Sequence 983, App
Sequence 2894, Ap
Sequence 425, App
Sequence 1301, Ap
Sequence 1301, Ap
Sequence 425, App
Sequence 45961, A
Sequence 45961, A
Sequence 45961, A
Sequence 11928, A
Sequence 11928, A
Sequence 13239, A
Sequence 13239, A
Sequence 32385, A
Sequence 32385, A
Sequence 32385, A
Sequence 1497, Ap
Sequence 42635, A
Sequence 1497, Ap
Sequence 21625, A
Sequence 2645, App
Sequence 314, App
Sequence 314, App
Sequence 19806, A
Sequence 16, Appl
Sequence 16, Appl
Sequence 16, Appl

32 201.4 9.4 572 16 US-09-205-070-7224 Sequence 7224, Ap
33 201.4 9.4 572 17 US-09-340-623-7224 Sequence 7224, Ap
34 201.4 9.4 572 33 US-09-898-888-7224 Sequence 7224, Ap
35 201.4 9.4 572 33 US-09-898-888A-7224 Sequence 7224, Ap
36 201.4 9.4 795 1 PCR-US01-08656-5196 Sequence 5196, Ap
37 200.2 9.4 352 19 US-09-515-128-15843 Sequence 15843, A
38 199.8 9.4 588 23 US-09-611-526-334 Sequence 334, App
39 199.8 9.4 588 23 US-09-611-526-1741 Sequence 1741, Ap
40 198.6 9.3 352 18 US-09-471-275-7634 Sequence 7634, Ap
41 193.8 9.1 377 27 US-09-699-999-4821 Sequence 4821, Ap
42 168 7.9 171 8 US-08-435-761-753 Sequence 753, App
43 147.4 6.9 1093 1 PCR-US01-08631-17949 Sequence 17949, A
44 135 6.3 418 33 US-09-866-555-10816 Sequence 10816, A
45 113 5.3 2984 55 US-60-167-217-12947 Sequence 12947, A

ALIGNMENTS

RESULT 1
US-09-743-237-3
; Sequence 3, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YUJI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743,237
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2134
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (407)...(1303)
US-09-743-237-3

Query Match 100.0%; Score 2134; DB 29; Length 2134;
Best Local Similarity 100.0%; Pred No. 0;
Matches 2134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 aattcgggtcaaggcgaagctcgcgggggcgacgacgagcgcggggagctcctcgggg 60
Db 1 aattcgggtcaaggcgaagctcgcgggggcgacgacgagcgcggggagctcctcgggg 60
QY 61 agtacccggggtccagagctcagcgctgagagagctcgagctcctcgagcccgccgc 120
Db 61 agtacccggggtccagagctcagcgctgagagagctcgagctcctcgagcccgccgc 120
QY 121 agcgcgccgctgcaagctgcaacttctctgctcgctgctacccgcgacgcgagccgc 180
Db 121 agcgcgccgctgcaagctgcaacttctctgctcgctgctacccgcgacgcgagccgc 180
QY 181 ggggttttgcctcgctggcgctggctcggaaggagctcccccggcgctccgc 240
Db 181 ggggttttgcctcgctggcgctggctcggaaggagctcccccggcgctccgc 240
QY 241 atgattccaggtgaatacaaggtaagcaggtggtactactacaagtaataatccggaaga 300
Db 241 atgattccaggtgaatacaaggtaagcaggtggtactactacaagtaataatccggaaga 300
QY 301 agcaactttgcaaatctcttctcaggaatctctgtgcaagttcccatggtcccgagg 360
Db 301 agcaactttgcaaatctcttctcaggaatctctgtgcaagttcccatggtcccgagg 360

QY 361 actagaggatgctcctcgtctgttcttcttaagaaagattccaccccaatggtgatatgcca 420
Db 361 actagaggatgctcctcgtctgttcttcttaagaaagattccaccccaatggtgatatgcca 420
QY 421 attgaaaggggacacacaaatctatgtatatacaaatctagaacaagagactaaaagc 480
Db 421 attgaaaggggacacacaaatctatgtatatacaaatctagaacaagagactaaaagc 480
QY 481 actccatttggctcctcagatatacaagatacaaaataattatctacagtcagatgcctaa 540
Db 481 actccatttggctcctcagatatacaagatacaaaataattatctacagtcagatgcctaa 540
QY 541 acaatgactgtttagtagtaggagatcttccagcatcaacaaataattatctacatctattac 600
Db 541 acaatgactgtttagtagtaggagatcttccagcatcaacaaataattatctacatctattac 600
QY 601 acaacaacttgaggagccttaccatcggtagtcaagggctgtcttccctcgggagtc 660
Db 601 acaacaacttgaggagccttaccatcggtagtcaagggctgtcttccctcgggagtc 660
QY 661 aactcttcaggagaccacccaaataaactttgctgggtactgtgactgtcttgcagtggtg 720
Db 661 aactcttcaggagaccacccaaataaactttgctgggtactgtgactgtcttgcagtggtg 720
QY 721 ggactttgcaacaactgcaatgttaataattgttgcaacaactgtcatcatgatatga 780
Db 721 ggactttgcaacaactgcaatgttaataattgttgcaacaactgtcatcatgatatga 780
QY 781 acggtttaaagccattaaagcagctcttctgtagaataccagagcttccagcacaataat 840
Db 781 acggtttaaagccattaaagcagctcttctgtagaataccagagcttccagcacaataat 840
QY 841 tgggaaggcccaatttgggcaatgtcaagcccgagcacaacaagggtgcaactgcaggag 900
Db 841 tgggaaggcccaatttgggcaatgtcaagcccgagcacaacaagggtgcaactgcaggag 900
QY 901 gtcaggtcctcgaagaataactgcagtgctatgagggcccaataatgtgtctctctat 960
Db 901 gtcaggtcctcgaagaataactgcagtgctatgagggcccaataatgtgtctctctat 960
QY 961 ttgcaaatgcattggtgcaaaaattatgaagaaagccagaaagcagaaacacataatgag 1020
Db 961 ttgcaaatgcattggtgcaaaaattatgaagaaagccagaaagcagaaacacataatgag 1020
QY 1021 catgccaactacatgcagactggaggttttggaggcagccattactgccaccacgaa 1080
Db 1021 catgccaactacatgcagactggaggttttggaggcagccattactgccaccacgaa 1080
QY 1081 atttcaggaactccaagattcaatcgaaggttggagggcagccattactgccaccacgaa 1140
Db 1081 atttcaggaactccaagattcaatcgaaggttggagggcagccattactgccaccacgaa 1140
QY 1141 ggttggtggaggccacatgcgctgcctgctgtctcagggagagagcgagagaaagaa 1200
Db 1141 ggttggtggaggccacatgcgctgcctgctgtctcagggagagagcgagagaaagaa 1200
QY 1201 ctgctccaagtgcctggcagagcagatgatccttgagggaatttggaggtgcttataca 1260
Db 1201 ctgctccaagtgcctggcagagcagatgatccttgagggaatttggaggtgcttataca 1260
QY 1261 gattctccacactgagtttaaatcgaaggttgaataatgagtagatataagtgga 1320
Db 1261 gattctccacactgagtttaaatcgaaggttgaataatgagtagatataagtgga 1320
QY 1321 atgcatgttgatttcttagcttagataatctctagttttagaagagatgttttaggggaa 1380
Db 1321 atgcatgttgatttcttagcttagataatctctagttttagaagagatgttttaggggaa 1380
QY 1381 catgaggtcgtcgtcagcaaacaccaggtcctcctgcatcctcgtcagggaggttt 1440
Db 1381 catgaggtcgtcgtcagcaaacaccaggtcctcctgcatcctcgtcagggaggttt 1440
QY 1441 actcagagctcctcgaagatgtggcaaccatgcccttctcgtcagggaggtgcatggcc 1500

Db	1590	TTAAATCTCATTTACACAACAACCTTGAGGAGAGCCTTACCATTGCGTAGTCAACGGGTCTGCT	1531
Qy	647	tccccctcgggatacactctccaggaccaccacaaaaataactttggctcgggtagctgtgac	706
Db	1530	TTCCCTTCGGGATCAACTCTTCCAGGACCACCNAAAAATACTTTGGCTGGGTACTGTGAC	1471
Qy	707	tgttttggcagtgggagacttttgcacaactgcaattgttaataattgttgcacaacttg	766
Db	1470	TGCTTTTGCAGATGGGACCTTTTGCACAACATGCAATTTGTAATAATTGTGTGCAACAACCTG	1411
Qy	767	catcatgatatagaacgggtttaaagccattaaagccattgaagcaatgtctgttagaataatcagaagact	826
Db	1410	CATCATGATATGTAACGGTTTAAAGCCATTAAAGGCATGCTTGTGTAGAAATCCAGAAGCT	1351
Qy	827	ttccagccaaaaatttggaagggccaaattgggcaatgtccaagccccagcacacaagaagg	886
Db	1350	TTCCAGCCAAAATTTGGGAAGGCCCAATTTGGCAATGTCAAGCCCCAGCACACAAGAGG	1291
Qy	887	tgcactgcaggaggtcaggtgcctcgaagaattactcgcagtgctatgaggcccaatt	946
Db	1290	TGCAACTGCAGGAGTCAAGCTGCCTGGAAGAAATTACTCGGAGTGTATGAGGCCCAATTT	1231
Qy	947	atgtgttctctatttgcaaatgcattggttgcaaaaattatgaagaaagcccgaaacga	1006
Db	1230	ATGTGTTCTTCATTGCAAAATGCAATTTGGTTGCAAAAAATTATGAAGAAAGCCCCAGAACA	1171
Qy	1007	aegacactaatgagcatgccaactcacatgacagactggaggtttggaagcgacgacattac	1066
Db	1170	AAGACACTAATGNGCATGCCAAACTTACATGCAGACTGGAGGTTTGGAAAGCGAGCCATTAC	1111
Qy	1067	ctgccaccacgaaattttcagggaattccaagattcacgcatagcggccttctctca	1126
Db	1110	CTGCCACCAACGAAATTTTCAGGACTTCCAAGATTCACTACGATAGCGGCGCTTCTCTCA	1051
Qy	1127	tgcattctctgggagtggtggaggccaacatgcgcctgcctgtctgcacgggagaagag	1186
Db	1050	TGCATCTCTGGGAGTGTGTGAGGCCACATGCGCTGCCCTGTGCTCAGGGGAAGAG	991
Qy	1187	gccgaaagaacactgctcccaagtgcctggcagagcagatgactcctggaggaattggga	1246
Db	990	GCCGAGAAAGAACACTGCTCCAAAGTGCCTGGCAGAGCAGATGATCTCTGGAGGAATTTGGA	931
Qy	1247	agtgcttatcacagattctccacactgagtttaactcgaaggattgaaaaatggagtag	1306
Db	930	AGGTGCTTTATCAGATTCTCCACACTGAGTTTAAATCTAAGGATGAAAAATGGAGTAG	871
Qy	1307	agataaagtgtgaatgcatgttgattttgtcttagtctagaataatctctagtttagaag	1366
Db	870	AGTATAAGTGTGAATGCATGTTGATTTGTGTTTATGCTAGAAATCTCTAGTTTAGAAG	811
Qy	1367	gagtgttagggaaacatgagctggctctgcagcaaacacagagctccccctgactccctg	1426
Db	810	GATGTTTAGGGGAACATGAGGCTGGCTCTGCAGCAACAACAGGCTCCCTGCATCCCTG	751
Qy	1427	ggccccaggagtttactcagagctctctgaagatgtggcaacccatgccccctttctga	1486
Db	750	GGCCCCAGGGAGTTTACTCAGAGCTCTCTGAAGATGTGCAACCAACCATGCCCTTTTCTCGA	691
Qy	1487	ggagggtgcatggcctgagcatgtttgtctgcccagaggagagagcttgggttccccata	1546
Db	690	GGAGGTGCATGGCCTGAGCATGTTGTTTGTCTGGCCCCAGAGAGAGCTTGGGTTTCCCAT	631
Qy	1547	gtcctgggagagtgtctcagggcggc-----	1573
Db	630	GTCTCTGGGAGAGTGTCTGCAGGGCGCGGAGGGGACACAGGGGAGGGGACAGCA	571
Qy	1574	-----ggagggcagagcaggccctgcggagagctcac	1605
Db	570	GGCCCTGCGAAGGCGACAGCGGGCAGGGGAGGCGAGAGGAGGAGCTTCC	511
Qy	1606	tctggtcgactcttctctcagagaattgtctcttggaggctgtctctgtcatgaataacccct	1665
Db	510	TCGTGTCGACTCTTCTCTTCACAGAAATGTGCTCTGAGAGCTGTCTGCGATGAAGACCTT	451

Qy	1666	aatggtttcttctgtttttcaaatatttagaataaagttctcccgatgggctgttgt	1725
Db	450	AATGGTTCTCTGTGTGTTTTCAAATTAATTAGAAATAAGTTCTCCGGATGGCTGTGT	391
Qy	1726	gataceacttaaaatcctctagagaactactgaacacctaagattttctgtagcgttagat	1785
Db	390	GATACCACCTTAAATCTCTAGAGAACTACTGAACACCTAAAGATTTTCTGTAGCGTAGAT	331
Qy	1786	atttccccagagacacgcgaactctcagtccttcttaagcccccgagacgcagcgcaa	1845
Db	330	ATTTCCCCACAGGCGACGGAACGTCTCAGTCTTTCTTAGGCCCCCGGAGACGCGACCA	271
Qy	1846	tggggccctcagcagcagcgttcgaccacgacatgtctttagttagtagagccttaaatatc	1905
Db	270	TGGGGCCTCGCAGCGCCAGCGTTTCACACGACATGTCTTGAGTTAGAGACTTTAAAATTATC	211
Qy	1906	cagttctctctgttcttacttgaattgtgaaaagcctctattccaatctaaactctc	1965
Db	210	CAGTTCTCTCTGTGTTTCTACTTTGAATTTGGAAAAGCTCTATTATCCAAATACTTCTC	151
Qy	1966	cataattattgttgtaatatattattgtttgtaaaacatggttcacataactagcttgt	2025
Db	150	CATAAATTATTGTTGTAATATTATTATTGTTTGTAAAACATGGTTCACATAACTAGCTTGT	91
Qy	2026	gaaacaccagcaggtataaagaattcctaagttgacgcgttttggtctctgtttaagcaaa	2085
Db	90	GGAACACAGCAGGTTAAATGAATCTTAAAGTTGACGCTTTTGGTTCTGTGTTAAAGCAA	31
Qy	2086	gatgaataaaaatttccaatgtcgaaaaa	2115
Db	30	GATCAATAAAAATTTCCAATGCTTCAAAA	1
RESULT 4			
US-09-743-237-2			
; Sequence 2, Application US/09743237			
; GENERAL INFORMATION:			
; APPLICANT: SUGIHARA, TAKASHI			
; APPLICANT: WADHWA, RENU			
; APPLICANT: KAUL, SUNIL C.			
; APPLICANT: MITSUI, YUJI			
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR			
; FILE REFERENCE: 084335/0127			
; CURRENT APPLICATION NUMBER: US/09/743,237			
; CURRENT FILING DATE: 2001-06-04			
; PRIOR APPLICATION NUMBER: PCT/JP99/03859			
; PRIOR FILING DATE: 1999-07-16			
; NUMBER OF SEQ ID NOS: 24			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 2			
; LENGTH: 1861			
; TYPE: DNA			
; ORGANISM: Mus musculus			
; FEATURE:			
; NAME/KEY: CDS			
; LOCATION: (271)..(1155)			
US-09-743-237-2			

[illegible]

167 cagcttcagagccctctggtcaggaatctgttgcaagttcccatccaggagg 226
363 tagagatgccctctgtcttcttaagaagaatcccaacaaatgggtatgccaat 422
227 cagagagccctcagctccttcggaagaagactccagccctcagctcagc 286
423 tgaagggggcacacaaatctatgatagacaattctagaacaagagaaacac 482
287 tgaagagggcccgagatgctctgatagacaactgtgagcagagagcacaagc 346
483 tccatttggtctcagatcagatcagaatcaaaataattatcagtcagatgctcctaaca 542
347 tccatctgctctcagatcagatcagagagagtttccctcagtcagagcctcaagc 406
543 caatgactcttagtaggagattttccagagcacaacaaattaaatctcattacac 602
407 caatgacaacttagtgggaagactctccagagacaggaattaaatctcattacac 466
603 acaacttgaggagccttaccatcgttagtcaacgggtctgttccctcgggacaa 662
467 aggtataatgagctcctccatcagctgcaatgggctgcttccctcgtgacctg 526
663 ctcttcagagaccacaaataaactttgggtggtgactgtgactgctttgcaagtgagg 722
527 ctctgcaagggccacccaaataaactctgtctggtgactgtgactgcttccagcggg 586
723 actttgcaacaaactcaattgaataattgttgcaacaaactgcatgatattgaac 782
587 actctgcaacagctgca-----gctgcaacaaactgcccagctgagctcagc 634
783 ggtttaaagcattaaagcagctgcttggtagaataccagaagctttccagccaaattg 842
635 gcttcaagcctaagaagcgtctgtgtagaatactgaaagctttccacccaaatgg 694
843 ggaagggccaattgggcaatgtcaagcccccagcaacaacaaagggtgcaactgcaggagt 902
695 ggaagggcgtctggagctgtcaaaacttcgacacagcaaggggtgcaactgtaagcgt 754
903 caggtctgctgaagaattactgcagtgctgactgagcccaaatatgtgtctctattt 962
755 caggtctgctgaagaactactgagtgctgactgagggcccaaatatgtgtctctattt 814
963 gcaaatgactgtgtcacaataattgaagaagccagaaacacgaacacaaatgagca 1022
815 gcaaatgactgtgtcacaataattgaagaagccagaaacacgaacacaaatgagca 874
1023 tgccaaactacatgcagactggaggtttggaagggcagcattaccctgcccacacgaat 1082
875 caccctacatggagcctgggagctttgagagcagcgaattattgtcccccagcgaat 934
1083 ttccaggaacttcaagattcagtcagatagcgccgtctctcctcatgcatctcctggagg 1142
935 tctcaggaacttcaaaactgagaaaaataggcagcctctcctgtatctcctgggaag 994
1143 tggtagggccacatgcgctgctgtgctcagggagagagggccgagaaacacact 1202
995 tagtggagggccacatgctgctgctgctgcccagggtaggaagcagagcagagcact 1054
1203 gctccaaagtgcctggcagacagatgactgaggaagatttgaaggtgcttatcaga 1262
1055 gttcccaagcttggtgagcagatgactgagggagtttgaaggtgctctgcaga 1114
1263 ttcccaactgagtttcaaatcagaaggttgaataaggagtagataaagtgtgaa 1322
1115 ttcccaactgagtttcaagtcacaaggggtgaaaaattgagtagcgtgcaagctggtaaa 1174
1323 gcatgttgattgtc 1338
1175 ggggaatgctgtggc 1190

US-09-743-237-1
; Sequence 1, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YUJI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743,237
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2241
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (651)..(1535)
US-09-743-237-1

Query Match 30.4%; Score 647.8; DB 29; Length 2241;
Best Local Similarity 76.1%; Pred. No. 2.le-94;
Matches 829; Conservative 0; Mismatches 247; Indels 13; Gaps 2;
QY 250 gttgaaatcaggttaagcaggtggtactactacaagtaataatccggaagaagcaacttt 309
Db 495 gttgaaatcaggttaagcaggtggtactactacaagtaataatccggaagaagcaacttt 553
QY 310 gcagaatctcttgcctcaggaatcctgttgcaggttcccatggtcccggaagactagagga 369
Db 554 ccaggccctctggtcaggaatcctgttgcaggttcccatggtcccggaagactagagga 613
QY 370 tgcctcctgctgtctcttaagaagaatcccaacaaatgggtgatatgcaattgaaagg 429
Db 614 ggctccagctgcctcgggaagaagactccagcccatgggtgattgttcagctgaaagg 673
QY 430 gggcacacaaatgctatgtagacaattctagaacaagactaaagcactcacttt 489
Db 674 agggcccgagatgctctgcatgagacaactgtgagcagggagagcctcactct 733
QY 490 ggttcctcagtatcaagatcaaaataattatctacagtcagatgctccctaaacacatgac 549
Db 734 gcttcctcagtatcaagatcaaaataattatctacagtcagatgctccctaaagcactgac 793
QY 550 tgccttagtagggagattttgcccagcatcaacaaataataatctcattacacacaaact 609
Db 794 aacttagtgggaagactctgcccagtcagcgaagttaaaatctcattacacacaggttga 853
QY 610 tggaggagccttaccatcagtagtcacaggtgctgttccctcgggatacactcttcc 669
Db 854 taatgagctctccatcagtagtcacaggtgctgttccctcgggactgctgca 913
QY 670 aggaccacacaaataaactttgggtgactgtgactgtgactgttggcagtgaggactttg 729
Db 914 agggccacacaaataaactctgctgggtgactgtgactgttccctcagcgggacttctg 973
QY 730 caacacactgcaattgtaataattgttgcaacaaactgcatcatgatatgaaagggttaa 789
Db 974 caacagctgca-----gctgcaacaaactgcatgagctgagcgtctcaa 1021
QY 790 agccattaaagcagctgttggtagaataccagagctttccagcccaaaaaattgggaagg 849
Db 1022 agccataaaggcgtgcttgatagaaatcctgaagctttccacacaaaaattgggaagg 1081
QY 850 ccaattgggcaatgtcaagccccagcacaacaaagggtgcaactgcagggagcgtg 909
Db 1082 ccgtctgggagctgctaaacttcgacacagcaagggtgcaactgtaagcgtcagcgtg 1141

Qy	910	cctgaagaataactgcgagtgctatgaggcccaaatattgtgttcttctatttgcgaatg	969
Db	1142	cctgaagaactactgtgagtgctatgaggcccaaatcatgtgttcttccatttgcgaatg	1201
Qy	970	cattggtgtcaaaaattatgaagaagccagacgaagaacactaatgagcatgcca	1029
Db	1202	cattggtgtcaaaaactatgaagaagtcagaaacgaagaaatgctgtatgagcacaccca	1361
Qy	1030	ctacatgcagactgaggtttggaagcagcattacctgccacaaacgaatttccagg	1089
Db	1262	ctacatgagcctgggactttgagagcagccattattgtccccagcgaattctcagg	1321
Qy	1090	acttccaagattcaatcacagataggcggccttctcctcatgcctctctgggaagtggtga	1149
Db	1322	acctccaacatgagaaaaataggcaggcctctccctgtatctcccgggaagtagtga	1381
Qy	1150	ggccacatgcctgcctgccttctcctcaggaggaagagcccgagaaagaaactgtccaa	1209
Db	1382	ggccacatgctcgtcgtcgtgccagggttgagaaagcagagcagacgtgttccc	1441
Qy	1210	gtgcctggagacagatgatcctcgtggagaaatttggaggtgtcttatcacagattctcca	1269
Db	1442	aagcttgcctgagcagatgatcctgagaggttttggaggtgcctgtcgagattctcca	1501
Qy	1270	cactgagatttaactcaagggattgaaaatggatagagtataaagtgtgaatgcagt	1329
Db	1502	catcgagttcaagtcacaaggggctgaaaattgagtagcgtgcgaagctggttaaagggaa	1561
Qy	1330	gattttgtc	1338
Db	1562	qcctgtatgc	1570

RESULT 6

```

RESULTS
US-60-360-207-983
; Sequence 983, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 983
; LENGTH: 2201
; TYPE: DNA
; ORGANISM: HUMAN
US-60-360-207-983

```

Query Match	29.7%	Score 633.8;	DB 75;	Length 2201;
Best Local Similarity	75.8%;	Pred. No. 3.7e-92;		
Matches 829;	Conservative	0;	Mismatches 247;	Indels 17;
				Gaps 3;

Qy	250	gttgaataatcaagtgaaagcagggtgggtactactacaagtaataatccgggaagaagcaacttt	309
Db	478	gttgaataatcaag-aagcagggtggtgagtgccaggcgccgagccctgaaagcgcagcgttt	536
Qy	310	gcagaaactcttgcctcagaaatccctgttgcaggttcccatggtccagagaaactagagg	369
Db	537	ccaggccctctggctcaggaaacctgtgtgcaagttcccatcatcccgaggcgagcaggga	596
Qy	370	tgcctctgtgttctcttaagaaaagtatccaaccctaatgggtgatgccaatttgaaagg	429
Db	597	ggctccagctgacctcggaagaagactccagcccatgggtatttgcagctgaaagg	656
Qy	430	gggcacacaaatgctatgtatagcaaatctagaaacagagaactaaagcactccatt	489
Db	657	aggcgccagatgctctgcatagacaactgtggcgcgaggaggctcaagcgtccatct	716
Qy	490	ggttcctcagtaatcaagatcaaaataattatctacagtcagatgtccctaaaccatgac	549
Db	717	gcttctcagtagacagacagcaggtttccctcaqtacagactccctcaagccaatgac	776

[illegible]

RESULTS

RESULTS /
US-09-611-526-2894
; Sequence 2894, Application US/09611526
; GENERAL INFORMATION:
; APPLICANT: OTA, TOSHIO
; APPLICANT: NISHIKAWA, TETSUO
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: HAYASHI, KOJI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: KAWAI, YURI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: KOJIMA, SHINICHI
; APPLICANT: OTSUJIMA, TETSUJI

; APPLICANT: KOGA, HISASHI
 ; TITLE OF INVENTION: PRIMERS FOR SYNTHESIS OF FULL LENGTH CDNAS
 ; FILE REFERENCE: 08335/0122
 ; CURRENT APPLICATION NUMBER: US/09/611,526
 ; PRIOR FILING DATE: 2000-07-07
 ; PRIOR APPLICATION NUMBER: JP 1999-194486
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: JP 2000-118774
 ; PRIOR FILING DATE: 2000-01-11
 ; PRIOR APPLICATION NUMBER: JP 2000-183765
 ; PRIOR FILING DATE: 2000-05-02
 ; NUMBER OF SEQ ID NOS: 4484
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2894
 ; LENGTH: 3495
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (88)..(840)
 ; US-09-611-526-2894

Query Match 22.3%; Score 476.8; DB 23; Length 3495;
 Best Local Similarity 87.1%; Pred. No. 7e-67;
 Matches 603; Conservative 0; Mismatches 7; Indels 82; Gaps 4;

QY	8	ggtcaaggcggaagctcgcgggggacagcagcgagcggggagctcctcggggagtaacc	67
DB	321	ggtcaaggcggaagctcgcgggggacagcagcgagcggggagctcctcggggagtaacc	380
QY	68	cgggattccagagctcagcgcgtgagagcgtcgcgtcctcagcgcgcgcgcgcgcgc	127
DB	381	cgggattccagagctcagcgcgtgagagcgtcgcgtcctcagcgcgcgcgcgcgcgc	440
QY	128	cgcctgcaacgtgcaactcctcgtcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtt	187
DB	441	cgcctgcaacgtgcaactcctcgtcctcgtcgtcgtcgtcgtcgtcgtcgtcgtt	500
QY	188	ttgcctcctgggcgtggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc	247
DB	501	---gcccctgggcgtggtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc	556
QY	248	cagtgtaaatcaaggtgaagcagtggtggtcgtcgtcgtcgtcgtcgtcgtcgtcgtc	307
DB	557	cagtgtaaatcaagg-aagcagtggtggtcgtcgtcgtcgtcgtcgtcgtcgtcgtc	615
QY	308	ttgcagaatctcttgcctcaggaatcctgtgtcgaagtcccatggtccctcaggaactagag	367
DB	616	ttgcagaatctcttgcctcaggaatcctgtgtcgaagtcccatggtccctcaggaactagag	675
QY	368	gatgcctcctggtctccttaagaagattcccaacccaatggtgatatgccaattgaaa	427
DB	676	gatgcctcctggtctccttaagaagattcccaacccaatggtgatatgccaattgaaa	735
QY	428	gggggcacacaaatgctatgtatagacaattctagacaagaagaactaaaagcactccat	487
DB	736	gggggcacacaaatgctatgtatagacaattctagacaagaagaactaaaagcactccat	795
QY	488	ttggttctcagtcatacaagatcaaaaataattatctacagtcagatgcctcctaaacaaatg	547
DB	796	ttggttctcagtcatacaagatcaaaaataattatctacagtcag- - - - -	838
QY	548	actgcttagtagggagatttttgcagcatcaacaaaataaattcattcattacacaaa	607
DB	839	-----	838
QY	608	cttgaggagccttacatcgttagtcaacgggtgtcgttcttccctcgggagtaactctt	667
DB	839	cttgaggagccttacatcgttagtcaacgggtgtcgttcttccctcgggagtaactctt	898
QY	668	ccaggaccaccacaaaataactttggtggtgta	699

Db 899 ccaggaccaccacaaaataactttggctgggta 930
 RESULT 8
 PCT-US01-01329-425
 ; Sequence 425, Application PC/TUS0101329
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences, Inc., et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PAI20PCT
 ; CURRENT APPLICATION NUMBER: PCT/US01/01329
 ; CURRENT FILING DATE: 2001-01-17
 ; NUMBER OF SEQ ID NOS: 3506
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 425
 ; LENGTH: 445
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (431)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (438)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; PCT-US01-01329-425

Query Match 20.1%; Score 429; DB 1; Length 445;
 Best Local Similarity 98.9%; Pred. No. 2.6e-59;
 Matches 440; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY	228	cccgggcgtcccgatgatcccgatgaaatcaaggtgaagcaggtggtactactacaagta	287
DB	2	cccgggcgtcccgatgatcccgatgaaatcaagg-aagcaggtggtactactacaagta	60
QY	288	ataatccggaaagaagaactttgcagaatctcttgcaggaatctcttgcaggaatctgtgcaggttcc	347
DB	61	ataatccggaaagaagaactttgcagaatctcttgcaggaatctcttgcaggaatctgtgcaggttcc	120
QY	348	catggtcccgaggaactagaggatgcctcgtgctgtcttcttaagaagaattccacccaa	407
DB	121	catggtcccgaggaactagaggatgcctcgtgctgtcttcttaagaagaattccacccaa	180
QY	408	tggatgatgccaatgaaagggggcacacaaatgctatgtatagacaattctagacaa	467
DB	181	tggatgatgccaatgaaagggggcacacaaatgctatgtatagacaattctagacaa	240
QY	468	gagaactaaaagcactccatttggctcctcagtcataagaatcaaaaataattctacagt	527
DB	241	gagaactaaaagcactccatttggctcctcagtcataagaatcaaaaataattctacagt	300
QY	528	cagatgtccctaaacaaatgactgcttttagtgaggagattttgcagcatcaacaaaat	587
DB	301	cagatgtccctaaacaaatgactgcttttagtgaggagattttgcagcatcaacaaaat	360
QY	588	taaatctattacacaaacttgaggagccttaccatcggtagtcacacgggtctgctt	647
DB	361	taaatctattacacaaacttgaggagccttaccatcggtagtcacacgggtctgctt	420
QY	648	tcctcctgggagtaactcttccagg 672	
DB	421	tcctcctgggagtaactcttccagg 445	

RESULT 9
 PCT-US01-01339-1301
 ; Sequence 1301, Application PC/TUS0101339
 ; GENERAL INFORMATION:
 ; APPLICANT: Human Genome Sciences, Inc., et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006PCT

; CURRENT APPLICATION NUMBER: PCT/US01/01339
 ; CURRENT FILING DATE: 2001-03-17
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1301
 ; LENGTH: 445
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE
 ; LOCATION: (431)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (438)
 ; OTHER INFORMATION: n equals a,t,g, or c
 PCT-US01-01339-1301

Query Match 20.1%; Score 429; DB 1; Length 445;
 Best Local Similarity 98.9%; Pred. No. 2.6e-59;
 Matches 440; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
 QY 228 ccggggcgctccgcatgacccagttgaatcaaggtaagcaggtggtactactacaagta 287
 Db 2 ccggggcgctccgcatgacccagttgaatcaagg-aagcaggtggtactactacaagta 60
 QY 288 ataaccggaagaagcaactttgcagaatcttctgctcaggaatcctgttgcagttcc 347
 Db 61 ataaccggaagaagcaactttgcagaatcttctgctcaggaatcctgttgcagttcc 120
 QY 348 catggtcccgaggaactagagatgcctcctgctgttcttgaagaagattcccaaccaa 407
 Db 121 catggtcccgaggaactagagatgcctcctgctgttcttgaagaagattcccaaccaa 180
 QY 408 tgggtatgatacgaattgaaagggggacacaaatgctatgtagacaattcttagaacia 467
 Db 181 tgggtatgatacgaattgaaagggggacacaaatgctatgtagacaattcttagaacia 240
 QY 468 gagaactaaagcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 527
 Db 241 gagaactaaagcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 300
 QY 528 cagatgctccctaaacaaatgactgctttagtagggagatttttgcagcatcaacaaaat 587
 Db 301 cagatgctccctaaacaaatgactgctttagtagggagatttttgcagcatcaacaaaat 360
 QY 588 taaatctcattacacaaacttgaggagccttaccatcggtagtcacacgggtctgctt 647
 Db 361 taaatctcattacacaaacttgaggagccttaccatcggtagtcacacgggtctgctt 420
 QY 648 tccctcgggagcaactcttccagg 672
 Db 421 tccctcgggntcaactnttccagg 445

RESULT 10
 US-09-764-891-1301
 ; Sequence 1301, Application US/09764891
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: PC006
 ; CURRENT APPLICATION NUMBER: US/09/764,891
 ; CURRENT FILING DATE: 2001-01-17
 ; Prior application data removed - consult PALM or file wrapper
 ; NUMBER OF SEQ ID NOS: 10231
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1301
 ; LENGTH: 445
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: SITE

; LOCATION: (431)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: SITE
 ; LOCATION: (438)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-09-764-891-1301
 Query Match 20.1%; Score 429; DB 30; Length 445;
 Best Local Similarity 98.9%; Pred. No. 2.6e-59;
 Matches 440; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
 QY 228 ccggggcgctccgcatgacccagttgaatcaaggtaagcaggtggtactactacaagta 287
 Db 2 ccggggcgctccgcatgacccagttgaatcaagg-aagcaggtggtactactacaagta 60
 QY 288 ataaccggaagaagcaactttgcagaatcttctgctcaggaatcctgttgcagttcc 347
 Db 61 ataaccggaagaagcaactttgcagaatcttctgctcaggaatcctgttgcagttcc 120
 QY 348 catggtcccgaggaactagagatgcctcctgctgttcttgaagaagattcccaaccaa 407
 Db 121 catggtcccgaggaactagagatgcctcctgctgttcttgaagaagattcccaaccaa 180
 QY 408 tgggtatgatacgaattgaaagggggacacaaatgctatgtagacaattcttagaacia 467
 Db 181 tgggtatgatacgaattgaaagggggacacaaatgctatgtagacaattcttagaacia 240
 QY 468 gagaactaaagcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 527
 Db 241 gagaactaaagcactcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 300
 QY 528 cagatgctccctaaacaaatgactgctttagtagggagatttttgcagcatcaacaaaat 587
 Db 301 cagatgctccctaaacaaatgactgctttagtagggagatttttgcagcatcaacaaaat 360
 QY 588 taaatctcattacacaaacttgaggagccttaccatcggtagtcacacgggtctgctt 647
 Db 361 taaatctcattacacaaacttgaggagccttaccatcggtagtcacacgggtctgctt 420
 QY 648 tccctcgggagcaactcttccagg 672
 Db 421 tccctcgggntcaactnttccagg 445

RESULT 11
 US-10-080-090-425
 ; Sequence 425, Application US/10080090
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 ; FILE REFERENCE: P120C1
 ; CURRENT APPLICATION NUMBER: US/10/080,090
 ; CURRENT FILING DATE: 2002-02-22
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 3506
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 425
 ; LENGTH: 445
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (431)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc_feature
 ; LOCATION: (438)
 ; OTHER INFORMATION: n equals a,t,g, or c
 US-10-080-090-425
 Query Match 20.1%; Score 429; DB 37; Length 445;
 Best Local Similarity 98.9%; Pred. No. 2.6e-59;

```
Matches 440; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
QY 228 ccgagcgcccgatgacccagtgaaatcaaggtgaagcaggtggtactactacaagta 287
Db 2 ccgagcgcccgatgacccagtgaaatcaaggtgaagcaggtggtactactacaagta 60
QY 288 ataatccggaagaagaacatttgcagaattcttctgctcaggaaatcctgttcaaggttcc 347
Db 61 ataatccggaagaagaacatttgcagaattcttctgctcaggaaatcctgttcaaggttcc 120
QY 348 catgtgccaggaactagaggatgctctgtgttcttcttaagaagaattccaacccaa 407
Db 121 catgtgccaggaactagaggatgctctgtgttcttcttaagaagaattccaacccaa 180
QY 408 tggatgatagcgaattgaaagggggcacacaatgctatgtatagacaattctagaacaa 467
Db 181 tggatgatagcgaattgaaagggggcacacaatgctatgtatagacaattctagaacra 240
QY 468 gagaactaaagcactccatttgggtctcctcagtatcaagatcaaaataattatctacagt 527
Db 241 gagaactaaagcactccatttgggtctcctcagtatcaagatcaaaataattatctacagt 300
QY 528 cagatgtccctaaacaaatgactgttttagtagggagatttttgcagcatcaacaaaat 587
Db 301 cagatgtccctaaacaaatgactgttttagtagggagatttttgcagcatcaacaaaat 360
QY 588 taaatctattacacaacacttgaggagcctaccatcgtagtcaacgggtctgtt 647
Db 361 taaatctattacacaacacttgaggagcctaccatcgtagtcaacgggtctgtt 420
QY 648 tccctcgggatacaactcttccagg 672
Db 421 tccctcgggatacaactcttccagg 445

RESULT 12
US-09-362-510-45961
; Sequence 45961, Application US/09362510
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-759CON1
; CURRENT APPLICATION NUMBER: US/09/362,510
; EARLIER FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 09/221,820
; NUMBER OF SEQ ID NOS: 62165
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45961
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(465)
; OTHER INFORMATION: n = A,T,C or G
US-09-362-510-45961

Query Match 18.9%; Score 404; DB 17; Length 465;
Best Local Similarity 98.3%; Pred. No. 2.7e-55;
Matches 415; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
QY 1150 ggccacatgcgctgctgtctgctcaggagaaagggccggaagaacactgctccaa 1209
Db 45 ggccacatgcgctgctgtctgctcaggagaaagggccggaagaacactgctccaa 104
QY 1210 gtgcttcgagacagatgactcctgagaaatttgaaggtgttatacagattctcca 1269
Db 105 -nnnnnngcagagcagatgactcctgagaaatttgaaggtgttatacagattctcca 163
QY 1270 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 1329
Db 164 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 223
QY 1330 gattttgtttagtctagaatctcttagtttagaagagatgttttaggggaacatagagct 1389
Db 224 gattttgtttagtctagaatctcttagtttagaagagatgttttaggggaacatagagct 283
QY 1390 gctctgcagacaacacagcgtccctgcattccctggccaggaggagtttactcagagc 1449
Db 284 gctctgcagacaacacagcgtccctgcattccctggccaggaggagtttactcagagc 343
```

```
Db 164 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 223
QY 1330 gattttgtttagtctagaatctcttagtttagaagagatgttttaggggaacatagagct 1389
Db 224 gattttgtttagtctagaatctcttagtttagaagagatgttttaggggaacatagagct 283
QY 1390 gctctgcagacaacacagcgtccctgcattccctggccaggaggagtttactcagagc 1449
Db 284 gctctgcagacaacacagcgtccctgcattccctggccaggaggagtttactcagagc 343
QY 1450 tctctgaagatgtggcaaccccatgccccctttcttgaggaggtgcattgagcctgagcattg 1509
Db 344 tctctgaagatgtggcaaccccatgccccctttcttgaggaggtgcattgagcctgagcattg 403
QY 1510 tttgtctggccagagagagacgttgggttcccatagtcctgggagagatgtctcaggg 1569
Db 404 tttgtctggccagagagagacgttgggttcccatagtcctgggagagatgtctcaggg 463
QY 1570 cg 1571
Db 464 cg 465

RESULT 13
US-09-362-510A-45961
; Sequence 45961, Application US/09362510A
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-759CON1
; CURRENT APPLICATION NUMBER: US/09/362,510A
; EARLIER FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 09/221,820
; NUMBER OF SEQ ID NOS: 62165
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45961
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(465)
; OTHER INFORMATION: n = A,T,C or G
US-09-362-510A-45961

Query Match 18.9%; Score 404; DB 17; Length 465;
Best Local Similarity 98.3%; Pred. No. 2.7e-55;
Matches 415; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
QY 1150 ggccacatgcgctgctgtctgctcaggagaaagggccggaagaacactgctccaa 1209
Db 45 ggccacatgcgctgctgtctgctcaggagaaagggccggaagaacactgctccaa 104
QY 1210 gtgcttcgagacagatgactcctgagaaatttgaaggtgttatacagattctcca 1269
Db 105 -nnnnnngcagagcagatgactcctgagaaatttgaaggtgttatacagattctcca 163
QY 1270 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 1329
Db 164 cactgagtttaaatctaaaggattgaaatggagtagagatataaaagtgtgaatgcatgtt 223
QY 1330 gattttgtttagtctagaatctcttagtttagaagagatgttttaggggaacatagagct 1389
Db 224 gattttgtttagtctagaatctcttagtttagaagagatgttttaggggaacatagagct 283
QY 1390 gctctgcagacaacacagcgtccctgcattccctggccaggaggagtttactcagagc 1449
Db 284 gctctgcagacaacacagcgtccctgcattccctggccaggaggagtttactcagagc 343
```

Qy 1450 tctctgaagatgtggcaacccatgcccccttttctgagagtgatgacctgagcattg 1509
|||||
Db 344 tctctgaagatgtggcaacccatgcccccttttctgagagtgatgacctgagcattg 403
|||||
Qy 1510 tttgtctgcccagagagagagcttggttcccatagctctctggagagtgctctgcagg 1569
|||||
Db 404 tttgtctgcccagagagagagcttggttcccatagctctctggagagtgctctgcagg 463
|||||
Qy 1570 cg 1571
||
Db 464 cg 465
||
RESULT 14
US-09-904-013-45961
; Sequence 45961, Application US/09904013
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-759CON1
; CURRENT APPLICATION NUMBER: US/09/904,013
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 09/362,510
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 09/221,820
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 62165
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45961
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(465)
; OTHER INFORMATION: n = A,T,C or G
US-09-904-013-45961

Query Match 18.9%; Score 404; DB 34; Length 465;
Best Local Similarity 98.3%; Pred. No. 2.7e-55;
Matches 415; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Qy 1150 ggcacatgcctgcctgcttgcctcaggagagagccgagagagcactgctccaa 1209
|||||
Db 45 ggcacatgcctgcctgcttgcctcaggagagagccgagagagcactgctccaa 104
|||||
Qy 1210 gtgctggcagagcagatgctcctggaggaatttggaggtgcttatcacagattctcca 1269
|||||
Db 105 -nnnnnngcagagcagatgctcctggaggaatttggaggtgcttatcacagattctcca 163
|||||
Qy 1270 cactgagtttaaatcagagagattgaaatggagtagagataaaagtgtgaatgcatgtt 1329
|||||
Db 164 cactgagtttaaatcagagagattgaaatggagtagagataaaagtgtgaatgcatgtt 223
|||||
Qy 1330 gattttgcttagctagaatactctagtttagaaagatgttttaggggaacatgagctt 1389
|||||
Db 224 gattttgcttagctagaatactctagtttagaaagatgttttaggggaacatgagctt 283
|||||
Qy 1390 ggcctgtagcaacaacacagagctcccctgcatcccctggcccgaggagtttactcagagc 1449
|||||
Db 284 ggcctgtagcaacaacacagagctcccctgcatcccctggcccgaggagtttactcagagc 343
|||||
Qy 1450 tctctgaagatgtggcaacccatgcccccttttctgagagtgatgacctgagcattg 1509
|||||
Db 344 tctctgaagatgtggcaacccatgcccccttttctgagagtgatgacctgagcattg 403
|||||
Qy 1510 tttgtctgcccagagagagagcttggttcccatagctctctggagagtgctctgcagg 1569
|||||
Db 404 tttgtctgcccagagagagagcttggttcccatagctctctggagagtgctctgcagg 463
|||||
Qy 1570 cg 1571

Db 464 cg 465
||
RESULT 15
US-08-971-197-11928
; Sequence 11928, Application US/08971197
; GENERAL INFORMATION:
; APPLICANT: Craig A. Rosen
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products - PO26
; NUMBER OF SEQUENCES: 12150
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,197
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PO26
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 11928:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-971-197-11928

Query Match 18.8%; Score 401.4; DB 13; Length 431;
Best Local Similarity 97.9%; Pred. No. 7.1e-55;
Matches 413; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
Qy 466 aagagaactaaagcactccattgtgttcctcagtatcaagatcaaaaataattatctaca 525
|||||
Db 8 AGGAGAACATAAAGCACCTCCATTTGGTTGCTCAGTATCAAGATCAAAATAATTATCTACA 67
|||||
Qy 526 gtccagatgcctcctaaaccaatgactgcttttagtggagagattttgccagcatcaacaaa 585
|||||
Db 68 GTCAGATGTCCTTAACCAATGACTGCTTTAGTAGGAGATTTTGCACGATCAACAAA 127
|||||
Qy 586 attaaatctcattacacaacaacttgaggagccttaccatcggttagtcaacgggtctgc 645
|||||
Db 128 ATTAATCTCATTACACAACAACACTTGAAGGAGCCCTTACCATCGGTAGTCAACGGGTCTGC 187
|||||
Qy 646 ttccctctcgagatcaactcttcaggaccacaaaataaactttggctgggtactgtga 705
|||||
Db 188 TTTCCCTTCGGGNTCACTNTTCCAGGCCACCAAAATAAATTTGGCTGGGTACTGTGA 247
|||||
Qy 706 ctgctttgccagtgaggacttttgcacaacactgcaattgtaataaattgttgcaacaact 765
|||||
Db 248 CTGCTTTCAGTGGGACTTTTGCAACANCTGCAATTGTAATAATTGTTGCANCAACTT 307
|||||
Qy 766 gcatcatgatattgaacgggttttaagccatttaagggatgctctgtagaataccagaagc 825
|||||
Db 308 GCATCATGATATTGAACGGTTTAAAGCCATTAAAGCCATTAAGGCATGCTCTTGGTAGAAATCCAGAGN 367
|||||

Search completed: July 11, 2002, 12:28:58
Job time: 15225 sec

QY 361 actagagatgctcctgctgttctcttaagaaagattccaaaccccaatggtgatgcca 420
 Db 361 actagagatgctcctgctgttctcttaagaaagattccaaaccccaatggtgatgcca 420
 QY 421 attgaaagggggcacacaaatgtatgtatagacaattctagacaagagaactaaagc 480
 Db 421 attgaaagggggcacacaaatgtatgtatagacaattctagacaagagaactaaagc 480
 QY 481 actccatttgggttccctcagtatcaagatacaaaataattatctcagtcagatgccctaa 540
 Db 481 actccatttgggttccctcagtatcaagatacaaaataattatctcagtcagatgccctaa 540
 QY 541 accaatgactgcttttagtagggagattttgcccagatcaacaaaattaaatctcattac 600
 Db 541 accaatgactgcttttagtagggagattttgcccagatcaacaaaattaaatctcattac 600
 QY 601 acaacaacttgaggggccttaccatggttagtcaacgggtctgctttccctcgggac 660
 Db 601 acaacaacttgaggggccttaccatggttagtcaacgggtctgctttccctcgggac 660
 QY 661 aactttccaggagcccaaaaataactttggtggtgactgtgactgctttccagtg 720
 Db 661 aactttccaggagcccaaaaataactttggtggtgactgtgactgctttccagtg 720
 QY 721 ggacttttgcacaactgcaattgttaataattgttgcacaacttgcacatgatattga 780
 Db 721 ggacttttgcacaactgcaattgttaataattgttgcacaacttgcacatgatattga 780
 QY 781 acgggttaagccattaaagcattgtcttggtagaataccagaagctttccagccaaaat 840
 Db 781 acgggttaagccattaaagcattgtcttggtagaataccagaagctttccagccaaaat 840
 QY 841 tgggaaggcccaattggcgaattgcaagcccgacacaaagggtgcacactgcaggag 900
 Db 841 tgggaaggcccaattggcgaattgcaagcccgacacaaagggtgcacactgcaggag 900
 QY 901 gtcaggctgctgaagaatttactgcagtgctatgagggcccaattatgtgtcttctat 960
 Db 901 gtcaggctgctgaagaatttactgcagtgctatgagggcccaattatgtgtcttctat 960
 QY 961 ttgcaaatgcatggttgcacaaattatgaagaagcccgagacgaagacactaatag 1020
 Db 961 ttgcaaatgcatggttgcacaaattatgaagaagcccgagacgaagacactaatag 1020
 QY 1021 catgccaaactacatgcagactgaggttggaggcagcattaccctgcacacacgaa 1080
 Db 1021 catgccaaactacatgcagactgaggttggaggcagcattaccctgcacacacgaa 1080
 QY 1081 atttccaggacttccagattcagatagcagtagcgcccttccatgcattctcctggga 1140
 Db 1081 atttccaggacttccagattcagatagcagtagcgcccttccatgcattctcctggga 1140
 QY 1141 ggtggtgagggccacatgcgcctgctgtctcaggggagagagcccgagaaagaa 1200
 Db 1141 ggtggtgagggccacatgcgcctgctgtctcaggggagagagcccgagaaagaa 1200
 QY 1201 ctgctcaagtgcctgagcagagatgctcctggaggaatttggaggtgcttatcaca 1260
 Db 1201 ctgctcaagtgcctgagcagagatgctcctggaggaatttggaggtgcttatcaca 1260
 QY 1261 gattctccacactgagtttaatacttaaggagattgaaatggagtagatgataagtgta 1320
 Db 1261 gattctccacactgagtttaatacttaaggagattgaaatggagtagatgataagtgta 1320
 QY 1321 atgcatgtgatttgccttagtagtaaaatctctagtttagaaagagatgttaggggaa 1380
 Db 1321 atgcatgtgatttgccttagtagtaaaatctctagtttagaaagagatgttaggggaa 1380
 QY 1381 catgaggtgactgctgcagcaaacacagagctccctgcacatccctggcccgaggagttt 1440
 Db 1381 catgaggtgactgctgcagcaaacacagagctccctgcacatccctggcccgaggagttt 1440

QY 1441 actcagagctctctgaagatgtggcaaccccatgcccccttttctgaggaggtgcagggcc 1500
 Db 1441 actcagagctctctgaagatgtggcaaccccatgcccccttttctgaggaggtgcagggcc 1500
 QY 1501 tgagcattgttctgtctgcccagagagagagagcttgggttcccatagtcctcggagagtg 1560
 Db 1501 tgagcattgttctgtctgcccagagagagagagcttgggttcccatagtcctcggagagtg 1560
 QY 1561 tctgcagggcggcgagagcagagcagggccctcgcgagagctcactctcgtgcactcttc 1620
 Db 1561 tctgcagggcggcgagagcagagcagggccctcgcgagagctcactctcgtgcactcttc 1620
 QY 1621 ctctcagaaatgttctctgagggtcgtctgcatagaacccctaatggttcttctgtt 1680
 Db 1621 ctctcagaaatgttctctgagggtcgtctgcatagaacccctaatggttcttctgtt 1680
 QY 1681 gtttttcaaatattttagaaaataagttctccggatgggtgttgcataccacttataat 1740
 Db 1681 gtttttcaaatattttagaaaataagttctccggatgggtgttgcataccacttataat 1740
 QY 1741 ctctagagaactactgaacacccctaaagatttctgtagcgtagatatttcccagagaca 1800
 Db 1741 ctctagagaactactgaacacccctaaagatttctgtagcgtagatatttcccagagaca 1800
 QY 1801 cgcgaactgtcagcttcttctaaggcccccggagagcagcagcgaatggggcctcgcaggc 1860
 Db 1801 cgcgaactgtcagcttcttctaaggcccccggagagcagcagcgaatggggcctcgcaggc 1860
 QY 1861 caggcttgaccagatgcttctgagttagtaggagacttaaaattatccagtttcttctgtg 1920
 Db 1861 caggcttgaccagatgcttctgagttagtaggagacttaaaattatccagtttcttctgtg 1920
 QY 1921 ttctacttgaattgtgaaaagctctattccaaatcttccatctcctcactaattattgtgt 1980
 Db 1921 ttctacttgaattgtgaaaagctctattccaaatcttccatctcctcactaattattgtgt 1980
 QY 1981 aatattattattgttgtgaaaacatggttccatcactgcttgggaaaccgagggta 2040
 Db 1981 aatattattattgttgtgaaaacatggttccatcactgcttgggaaaccgagggta 2040
 QY 2041 aaatgaattcttaagttgacgcttttgggttctgttgaagcaagatgaataaaattt 2100
 Db 2041 aaatgaattcttaagttgacgcttttgggttctgttgaagcaagatgaataaaattt 2100
 QY 2101 ccaatgtcgaaaaaataaaataaaataaaataaa 2134
 Db 2101 ccaatgtcgaaaaaataaaataaaataaaataaa 2134

RESULT 2

US-10-112-699-2790, Application US/10112699

GENERAL INFORMATION:
 APPLICANT: Wang, Tongtong
 APPLICANT: Wang, Siqing
 APPLICANT: Bangur, Chaitanya S.
 APPLICANT: Gaiger, Alexander
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF CANCER
 FILE REFERENCE: 210121.565
 CURRENT APPLICATION NUMBER: US/10/112,699
 CURRENT FILING DATE: 2002-03-28
 NUMBER OF SEQ ID NOS: 2959
 SEQ ID NO 2790
 LENGTH: 2134
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-10-112-699-2790

Query Match 99.7%; Score 2127.6; DB 6; Length 2134;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 2130; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	1	aattcggggtcaaggctcgcggggggcagacgacgagcgggggagtctctcctcggggg	60
Db	1	aattcggggtcaaggcgaagctcgggggggcagacgacgagcgggggagtctcctcgggg	60
Qy	61	agtaccccggatcccagaagctcagcgacgtctgagagcgtcgactcctcgagccccgc	120
Db	61	agtaccccggatacccagaagctcagcgacgtctgagagcgtcgactcctcgagccccgc	120
Qy	121	agccgccgcctgcaaccttcacttctctgctcgtctaccgcgcacgcagccccgc	180
Db	121	agccgccgcctgcaacgtgcacttccitgctcgtctaccgcgcacccgacccccgc	180
Qy	181	gggtgttttcccctgggcgcctgggtcctgcgaagagagcctcccaccocggcgctccg	240
Db	181	gggtgttttcccctgggcgcctgggtcctgcgaagagagcctcccaccocggcgctccg	240
Qy	241	atgatcccaagtgaatacaagtgtaagcaggtgggtactactacaagttaatcgcgaaga	300
Db	241	atgatcccaagtgaatacaagtgtaagcaggtgggtactactacaagttaatcgcgaaga	300
Qy	301	agcaactttgcagaatcttttgtctcgaagaatctctgtgtgcagttcccatgggtcccgga	360
Db	301	agcaactttgcagaatcttttgtctcaggaatctctgtgtgcagttcccatgggtcccgga	360
Qy	361	actagagatgcctcctgctgtcttctaagaaagattccaaccaaatgggtatatgccca	420
Db	361	actagagagtcctcctcgtgttctcttaagaaagattccaaccaaatgggtatatgccca	420
Qy	421	attgaagaggggcacacaaatgctatgtatagacaattctagaacaagaaactaaaagc	480
Db	421	attgaagaggggcacacaaatgctatgtatagacaattctagaacaagaaactaaaagc	480
Qy	481	actccatttggttcttcagta caagata caaaataaatttatctacgtcagatgtccctaa	540
Db	481	actccatttggttcttcagtatcaagata caaaataaatttatctacgtcagatgtccctaa	540
Qy	541	accaatgactgttttagttagggagattttggccagcatacaaaaaataaattctcattac	600
Db	541	accaatgactgtcttagttagggagattttggccagcatacaaaaaataaattctcattac	600
Qy	601	acaacaacttgagggagccttacca tgggtagtacaaggggtctgctttccccctgggattc	660
Db	601	acaacaacttgagggagccttaccatcggtagtacaaggggtctgctttccccctgggattc	660
Qy	661	aactcttcaggagaccaccaaaaataaactttggctgggtgactgtgactgctttggccagtgg	720
Db	661	aactcttcaggagaccaccaaaaataaactttggctgggtgactgtgactgctttggccagtgg	720
Qy	721	ggacttttgcacaactgcgaattgtaataattgtttgcacaacitgcatactgatattga	780
Db	721	ggacttttgcacaactgcgaattgtaataattgtttgcacaacitgcatactgatattga	780
Qy	781	acgggtttaagggcattaaaggcatgtctgtgtgtagaaatccagaagctttccagccaaaaat	840
Db	781	acgggtttaagggcattaaaggcatgtctgtgtgtagaaatccagaagctttccagccaaaaat	840
Qy	841	tgggaagggccaattgggccaatgtcaagccccagcacacaaaaggggtgcaactgcaggag	900
Db	841	tgggaagggccaattgggccaatgtcaagccccagcacacaaaaggggtgcaactgcaggag	900
Qy	901	gtcaggtcgtctgaagaatactcgcagtgctcatgagcccccaaatattggttcttctat	960
Db	901	gtcaggtcgtctgaagaatactcgcagtgctcatgagcccccaaatattggttcttctat	960
Qy	961	ttgc aaatgc attt ggttgc aaaaattat ga agaa gccc ga acgaa gac actaa tt gag	1020
Db	961	ttgc aaatgc attt ggttgc aaaaattat ga agaa gccc ga acgaa gac actaa tt gag	1020
Qy	1021	catgc aaaa ctac at gc ag act gg ag gttt ga ag ggc gcc acc tac ct gc ccc aac ga a	1080
Db	1021	catgc aaaa ctac at gc ag act gg ag gttt ga ag ggc gcc acc tac ct gc ccc aac ga a	1080

[illegible]

Db 241 ccaatgactgtcttagtaggagagatttttgcagacatcaacaaataaataatctcattaca 300
Qy 602 caacaactgaggagccttaccatcggtagtcacacgggtctgcttccctccgagatca 661
|||||
Db 301 caacaactgaggagccttaccatcggtagtcacacgggtctgcttccctccgagatca 360
Qy 662 actcttccaggaccaccacaaataaacttggctgggtactgtgactgtttgcccagtggg 721
|||||
Db 361 actcttccaggaccaccacaaataaacttggctgggtactgtgactgtttgcccagtggg 420
Qy 722 gactttgcaacaactgcaattgtaataattgttgcacaacacttgcacatgatattgaa 781
|||||
Db 421 gactttgcaacaactgcaattgtaataattgttgcacaacacttgcacatgatattgaa 480
Qy 782 cgggttt 787
|||||
Db 481 cggntt 486

RESULT 6
US-10-106-698-2655
; Sequence 2655, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: Patentin Ver. 3.0
; SEQ ID NO 2655
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (199)..(199)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (204)..(204)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (214)..(214)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (232)..(232)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (243)..(243)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-2655

Query Match 11.5%; Score 245; DB 6; Length 250;
Best Local Similarity 98.0%; Pred. No. 1.5e-37;
Matches 245; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1144 ggtgagagccacatgcctgcctgcttctcaggagagagcgcgagagaaacactg 1203
|||||
Db 1 ggtgagagccacatgcctgcctgcttctcaggagagagcgcgagagaaacactg 60
Qy 1204 ctccaagtgcctggcagacagatgacatccctggaggaatttgggaaggtgcttatcacagt 1263
|||||
Db 61 ctccaagtgcctggcagacagatgacatccctggaggaatttgggaaggtgcttatcacagt 120
Qy 1264 tctccacactgagtttaaatctaaaggagattgaaatggagtagagataaagtgtaag 1323
|||||

Db 121 tctccacactgagtttaaaatctaaaggattgaaaatggagtagagtataaaagtgtgaatg 180
Qy 1324 catgttgatttgccttagtctagaataatctctagtttagaagaagatgttttaggggaacat 1383
|||||
Db 181 catgttgatttgccttagtctagaataatctctagtttagaagaagatgttttaggggaacat 240
Qy 1384 gaggctggct 1393
|||||
Db 241 gaggctggct 250

RESULT 7
US-09-721-544-15843
; Sequence 15843, Application US/09721544
; GENERAL INFORMATION:
; APPLICANT: Arterburn, Matthew
; APPLICANT: Asghari, Vida
; APPLICANT: Damavandi, Simin
; APPLICANT: Dickson, Mark
; APPLICANT: Drake, Jim
; APPLICANT: Drmanac, Radoje
; APPLICANT: Engleman, Carrie
; APPLICANT: Faulkner, Brandy
; APPLICANT: Garcia, Veronica
; APPLICANT: Giedt, Gretchen
; APPLICANT: Hunter, Kelly
; APPLICANT: Jensen, Aaron
; APPLICANT: Jones, Lee
; APPLICANT: Kita, David
; APPLICANT: Labat, Ivan
; APPLICANT: Laroya, Mimi
; APPLICANT: Lomelli, Michelle
; APPLICANT: Nguyen, Phuong
; APPLICANT: Nogra, Margie
; APPLICANT: Palencia, Servando
; APPLICANT: Raisi, Fariba
; APPLICANT: Smith, Benjamin
; APPLICANT: Tkach, Joe
; APPLICANT: Tran, Lien
; APPLICANT: Verna, Ron
; APPLICANT: Yang, Fei
; APPLICANT: Yim, Kenneth
; TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
; TITLE OF INVENTION: From a CDNA Library of Fetal Liver-Spleen
; FILE REFERENCE: 728CIP
; CURRENT APPLICATION NUMBER: US/09/721,544
; CURRENT FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 09/515,128
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 09/034,341
; PRIOR FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 24489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15843
; LENGTH: 352
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(352)
; OTHER INFORMATION: n = A,T,C or G
US-09-721-544-15843

Query Match 9.4%; Score 200.2; DB 5; Length 352;
Best Local Similarity 75.1%; Pred. No. 5.2e-29;
Matches 250; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
Qy 697 gactgtgactgttgcagctgtgggacattttcaacaacactgcaattgtaataattgttg 756
|||||
Db 4 gattgtgattgttgcagaaatgtgaaattttgcaaacactgcaattgtactaatgttta 63
Qy 757 caacaactgcatcatgatattgaaagggttttaaaagccattaaagccatgctgttgtagaaa 816

Db 64 caacaatttggacatgaaataagaaagcaaaagcaataaaggcatgcttgcacagaa 123
Qy 817 tccagaagcttcagcccaaaaattggaaaggccaattgggcaatgtcagagcccaagca 876
Db 124 tccagaagctttaaagcttaagtagaggaaggaaggaaggaaggaaggaaggaagga 183
Qy 877 caacaaggggtgcaactgagggaggtcagctgctgctgaagaattactcgagtgctatga 936
Db 184 tagcaagggtgttaattgcaaacgcatcaggatgtcttaaaaactactgtgaatgctatga 243
Qy 937 ggcacaaattatgtgttcttatttgcacaaatgcatgctggttgcaaaaaattatgaagaag 996
Db 244 ggcacaaataatgttctcctaataatgcaaatgattgctgtaagaatttttgaagaag 303
Qy 997 cccagaagaagacactaatgagcatgccaaa 1029
Db 304 cccgaaaggaagacattgatgtctttggcaga 336

RESULT 8

US-10-155-881-3628

; Sequence 3628, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lutfiyya, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 3628

; LENGTH: 1684

; TYPE: DNA

; ORGANISM: Zea mays subsp. mexicana

US-10-155-881-3628

Query Match 4.4%; Score 93.6; DB 6; Length 1684;
Best Local Similarity 57.5%; Pred. No. 1.1e-08;
Matches 208; Conservative 0; Mismatches 124; Indels 30; Gaps 1;
Qy 683 ataaacttggctgggtactgtgactgtgttgcagtgagggaacttttgcacaaactgcaat 742
Db 241 aaatgttgaagtgtactgtgaatgttcttctgttacttactgtgtggttgcac 300
Qy 743 tgaataattgtgcacaaactgtcatgatattgaacggttttaaagccattgaagca 802
Db 301 tgcacaaactgttttaataactgtataggttgcaggcgtgagcgtgtgtaagct 360
Qy 803 tgtcttggtagaataccagaagcttccagcccaaaaattggaaaggccaattgggcaat 862
Db 361 atattggatcgaataccagatgttccagcctaagattggagcagccacttatgcat 420
Qy 863 gtcaagccccc-----agcacacaaaagggtgcac 892
Db 421 aggaataatgaggtgctgtagtgcctctttagtagggaagcacaataaagggtgccat 480
Qy 893 tcaagaggtcaggtgcgtcgaagattactgcagatgtcagatgagcccaattatgtgt 952
Db 481 tgcagaagagtcaggatgtctcaagaagctactgtgtgaggtcttccaaagcctaatactatgc 540
Qy 953 tcttctattgcaaatgattgttgcacaaaattatgaagaagcccaagcgaagca 1012
Db 541 tctgaaactgcagtgcatgattgcagaagaattttgaaggagtggaagagagaagaat 600
Qy 1013 ct 1014
Db 601 ct 602

RESULT 9

US-10-155-881-3943

; Sequence 3943, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lutfiyya, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 3943

; LENGTH: 1254

; TYPE: DNA

; ORGANISM: Zea mays

US-10-155-881-3943

Query Match 3.1%; Score 66.4; DB 6; Length 1254;
Best Local Similarity 65.5%; Pred. No. 0.0016;
Matches 97; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
Qy 863 gtcaagcccccagcacacaaagggtgcacactgcagaggtcagaggtcaggtgcctgaagaattac 922
Db 9 ggcggggcaagcataaacccaggtgtcactgcagaagtcaggtgtgtctaaagaataac 68
Qy 923 tgcagatgtctatgagcccaaatatgtgttcttcttcaaatgcattggttgcaaa 982
Db 69 tgcagatgttttcaagcgaatactccttgcctcaagaattgtcaaatgtatgattgcaaa 128
Qy 983 aattatgaagaagcccgacgaaga 1010
Db 129 aactatgaaggagtgaggagctacgga 156

RESULT 10

US-10-155-881-3945

; Sequence 3945, Application US/10155881

; GENERAL INFORMATION:

; APPLICANT: Dotson, Stanton B.

; APPLICANT: Kovalic, David K.

; APPLICANT: Liu, Jingdong

; APPLICANT: Lutfiyya, Linda L.

; APPLICANT: McIninch, James

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: US/10/155,881

; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595

; SEQ ID NO 3945

; LENGTH: 900

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)-(900)

; OTHER INFORMATION: unsure at all n locations

US-10-155-881-3945

Query Match 2.7%; Score 57; DB 6; Length 900;
Best Local Similarity 58.6%; Pred. No. 0.091;
Matches 99; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
Qy 682 aataacttggctgggtactgtgactgttgcagtgagggtggtgcaactgtcaa 741
Db 601 ct 602


```

Db 614 aaaaattgtgaagatgactgtgaatgtcttcaagaacttcaatactgtgaggtgcaa 673
Qy 742 ttgtaataattgtgcaacaactgtcatcatgatattgaacgttttaaagccattaaagc 801
Db 674 cftttcaaatgtggaacattgttgggaatgagaatgccagaacgaagaacttgagcc 733
Qy 802 atgtctgtgtagaataccagaagcttccagcccaaaaaattgggaagggc 850
Db 734 aattgcgcagtaataccgtctgctgtttccagctaaattgggaatgac 782

```

RESULT 11

```

US-10-027-632-221356
; Sequence 221356, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 221356
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-221356

```

```

Query Match 2.6%; Score 56; DB 6; Length 530;
Best Local Similarity 77.3%; Pred. No. 0.13;
Matches 68; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```

```

Qy 942 aaatatgttcttcttatttcaaatgcatgtgttgcaaaaattatgaagaagcccg 1001
Db 3 aaataatgttcttcttcaaatgcaaatgtattgtgtgaagaattttgaagaagcccg 62
Qy 1002 aacgaagaagactaatgagatgccaaa 1029
Db 63 aaagaagaagactgattgcatgttgccaga 90

```

RESULT 12

```

PCR-US02-01109-45
; Sequence 45, Application PC/TUS0201109
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: P2016PCT2
; CURRENT APPLICATION NUMBER: PC/TUS02/01109
; CURRENT FILING DATE: 2002-01-17
; PRIOR APPLICATION NUMBER: US 60/262,066
; PRIOR FILING DATE: 2001-01-18
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 986
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
PCR-US02-01109-45
Query Match 2.4%; Score 52; DB 1; Length 986;
Best Local Similarity 52.8%; Pred. No. 0.82;
Matches 112; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
Qy 1923 ctacttgaaattggaaaagctctattatcccaatttaacttccataattattgtgttaa 1982
Db 702 ctgacagatgtggaagaaggaatcagatcttcttctggtttatactggtgttaa 761
Qy 1983 tattattattgtttgtaaaacatggttcacataactagcttggaaaaccagcaggtaaa 2042
Db 762 aacagaatgatcacagaaaatgtttcttcttgaactgtagttgaacatagaactggg 821
Qy 2043 atgaattcttaagttagcgcgttttggcttctgtttaaagcaagaatgaataaaatttcc 2102
Db 822 tattatagatcaacttttcaacttttgggaatgtttgtattgaacttaataaaacttaa 881
Qy 2103 aatgtcgaaaaaataaaaaataaaaaaataaaaaa 2134
Db 882 catggcaaaaaaataaaaaaataaaaaaataaaaaa 913

```

RESULT 13

```

US-10-102-806-196
; Sequence 196, Application US/10102806
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA103PIC1
; CURRENT APPLICATION NUMBER: US/10/102,806
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: 09/925,298
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PC/US00/05881
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 846
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 196
; LENGTH: 1772
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1749)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1769)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-102-806-196

```

```

Query Match 2.4%; Score 52; DB 6; Length 1772;
Best Local Similarity 52.8%; Pred. No. 0.89;
Matches 112; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
Qy 1923 ctacttgaaattggaaaagctctattatcccaatttaacttccataattattgtgttaa 1982
Db 1473 ctgacagatgtggaagaaggaatcagatcttcttctggtttatactggtgttaa 1532
Qy 1983 tattattattgtttgtaaaacatggttcacataactagcttggaaaaccagcaggtaaa 2042
Db 1533 aacagaatgatcacagaaaatgtttcttcttgaactggttagttgaacatagaactggg 1592
Qy 2043 atgaattcttaagttagcgcgttttggcttctgtttaaagcaagaatgaataaaatttcc 2102

```


Db 1593 tattatagatcacttttcacttttttggaatgttttgattgaaacttaataaaaacttttaa 1652
QY 2103 aatgtcgaaaaaataaaaaaataaaaaa 2134
Db 1653 catggcaaaaaaataaaaaaataaaaaa 1684

RESULT 14

US-10-138-145-301
; Sequence 301, Application US/10138145
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: Gibson, John Bryan
; APPLICANT: Norris, Michael Geoffrey
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use.
; FILE REFERENCE: 11000.1058U
; CURRENT APPLICATION NUMBER: US/10/138,145
; CURRENT FILING DATE: 2002-05-01
; NUMBER OF SEQ ID NOS: 1652
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 301
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Lolium perenne
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(381)
; OTHER INFORMATION: n = A,T,C or G
US-10-138-145-301

Query Match 2.4%; Score 51.4; DB 6; Length 381;
Best Local Similarity 65.0%; Pred. NO. 0.94;
Matches 76; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 868 gccccagcacacaagaagggtgcaactgaggaggtcaggctgcctgcaagaataactgcga 927
Db 253 ggcctgctataagagagaggtgcaattgcaagaagtcgaagctgtctcaagaaatattgcga 312
QY 928 gtgctatgagggcccaaatgtgtctcttattgcaaaatgcattgttgcaaaaa 984
Db 313 atgctatcaggagggtgtgtgctgctcacaaactgcagatgcgaacttgcaaaa 369

RESULT 15

US-10-155-881-3101
; Sequence 3101, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 3101
; LENGTH: 763
; TYPE: DNA
; ORGANISM: Zea mays
US-10-155-881-3101

Query Match 2.3%; Score 49.4; DB 6; Length 763;
Best Local Similarity 66.4%; Pred. NO. 2.5;
Matches 71; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 878 aacaaagggtgcaactgcaggaggtcaggtcctgcctgaaataattactgcagtgctatgag 937
Db 599 aagagagggttgcactgcaggaagtcaggctgcttaagaaatactctgtgattgctaccag 658
QY 938 gcccaaatattgtttcttcttatttgcaaatgcattggttgcaaaaa 984
Db 659 gatggactggatgttctcttcttctgcccgtgtgacgattgccagaa 705

Search completed: July 11, 2002, 12:40:34
Job time: 12441 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:13:36 ; Search time 53.06 seconds
(without alignments)
617.342 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVTCLKGAQMLCIDNCGA.....GRLCSILHIEFKSLKLTIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1599	100.0	295	21	AA198463 Mouse testis speci
2	1251	78.2	299	21	AA198464 Human testis speci
3	435	27.2	950	22	AB862035 Drosophila melanog
4	396	24.8	438	20	AAW83392 Caenorhabditis ele
5	362	22.6	280	22	ABG17958 Novel human diagn
6	322	20.1	147	22	AAW95330 Human reproductive
7	248	15.5	53	22	AAW76158 Human colon cancer
8	186	11.6	243	22	AB868888 Drosophila melanog
9	165	10.3	251	22	AAW93348 Human polyptide,
10	110.5	6.9	732	18	AAW26642 Human RECK cancer-
11	99	6.2	2931	22	AB868229 Drosophila melanog

12	97	6.1	602	21	AA195660 Human Zntr2 protei
13	97	6.1	1987	22	AB861083 Drosophila melanog
14	96.5	6.0	359	21	AA194997 Arabidopsis thalia
15	96.5	6.0	443	21	AA194996 Arabidopsis thalia
16	96.5	6.0	472	21	AA194995 Arabidopsis thalia
17	96.5	6.0	1788	22	AB862995 Drosophila melanog
18	96.5	6.0	4547	22	AB859051 Drosophila melanog
19	96	6.0	773	22	AB859751 Drosophila melanog
20	95.5	6.0	1035	22	AB866062 Drosophila melanog
21	94	5.9	1792	21	AA194843 Human laminin 8 po
22	94	5.9	1800	21	AA194844 Human laminin 8 po
23	94	5.9	1816	21	AA194844 Human laminin 8 po
24	94	5.9	1824	21	AA194844 Human laminin 8 po
25	93.5	5.8	652	19	AA194879 Amino acid sequenc
26	93.5	5.8	652	21	AA194879 Human cell surface
27	93.5	5.8	1981	22	AB861657 Drosophila melanog
28	92.5	5.8	921	21	AA194297 Corn polycomb prot
29	92	5.8	3680	22	AB870878 Drosophila melanog
30	91.5	5.7	637	21	AA1942346 Mouse cell surface
31	91.5	5.7	644	21	AA1942346 Haematopoietic ste
32	91.5	5.7	644	21	AA1942346 Haematopoietic ste
33	91	5.7	18	21	AA1942346 Mouse testis speci
34	89.5	5.6	347	22	AA1942346 Human gene 35 enco
35	89.5	5.6	361	22	AA1942346 Human polyptide
36	89.5	5.6	439	22	AA1942346 Human protein sequ
37	89	5.6	1725	21	AA194800 Mouse laminin 2 ma
38	89	5.6	1725	21	AA194800 Mouse laminin 8 po
39	89	5.6	1764	10	AA1948451 Primary amino acid
40	89	5.6	1776	19	AA1948094 Mouse laminin B1 c
41	89	5.6	1786	21	AA1947999 Mouse laminin 2 be
42	89	5.6	1786	21	AA1948450 Mouse laminin 8 po
43	88.5	5.5	3843	22	AB871529 Drosophila melanog
44	88	5.5	250	21	AA1942335 Metallothionein ve
45	88	5.5	466	11	AA1947447 Human laminin B1 c

ALIGNMENTS

RESULT 1

AA1948463

ID AA1948463 standard; Protein: 295 AA.

AC AA1948463;

DT 25-APR-2000 (first entry)

DE Mouse testis specific factor tesmin SEQ ID NO:4.

Testis specific factor: tesmin; cell death; regulation; spermatocyte;

KW differentiation regulatory factor; male germ cell regulatory actor;

KW germ cell differentiation; sterility.

OS Mus musculus.

PN WO200004147-A1.

PD 27-JAN-2000.

PF 16-JUL-1999; 99WO-JP03859.

PR 17-JUL-1998; 98JP-0219856.

PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

PI Sugihara T, Wadawa R, Kaul SC, Mitsui Y;

DR WPI; 2000-147785/13.

DR N-PSDB; AA288155, AA288157.

PT New male germ cell regulatory factor tesmin expressed in spermatocytes
PT useful for investigation of germ cell differentiation and sterility -

XX Claim 1; Page 47-49; 63pp; Japanese.
 PS The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin.
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX
 SQ Sequence 295 AA;

Query Match 100.0%; Score 1599; DB 21; Length 295;
 Best Local Similarity 100.0%; Pred. No. 2.5e-144; Indels 0; Gaps 0;
 Matches 295; Conservative 0; Mismatches 0;

QY 1 MVTCLKGGAGMLCNDCCARELKHLLPOYDDQSFPQSELKPKMTTLVGRLLPVPK 60
 DB 1 mvtclkggagmlcndccarelkahlhlpqyddqsfpqselkpkmttlvgrllvpak 60
 QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTITLGGYDCFSGDFCNCSNLRHEL 120
 DB 61 lnlitqvdngalpsavngaappspalqgpptitlsgydcfsgdfcnscnnlrhel 120
 QY 121 ERFAKACLDNRNPEAFQPMKGRGRLGAALKRHSKGCNCRSGCLKNYCEYEAQIMCSS 180
 DB 121 erfakacldnrnpeafqpmkgrgrrlgaalkrshskgcnckrsgclknycyeyeaqimcss 180
 QY 181 IKCICKACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISW 240
 DB 181 icckickacknyeesperkmlmstphymepgdfesshylsapakfsppklrknrqafscisw 240
 QY 241 EVVEATCACLQAQEEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
 DB 241 evveatcaccllaqeeaeqhcspslaemileefgrclsqilhieffkskglkme 295

RESULT 2
 AAY68464
 ID AAY68464 standard; Protein; 299 AA.
 AC AAY68464;
 XX
 XX 25-APR-2000 (first entry)
 DT
 DE Human testis specific factor tesmin SEQ ID NO:5.
 XX
 XX Testis specific factor; tesmin; cell death; regulation; spermatocyte;
 KW differentiation regulatory factor; male germ cell regulatory actor;
 KW germ cell differentiation; sterility.
 XX
 OS Homo sapiens.
 XX
 XX WO200004147-A1.
 PN
 XX 27-JAN-2000.
 PD
 XX
 XX 16-JUL-1999; 99WO-JP03859.
 PF
 XX
 XX 17-JUL-1998; 98JP-0219856.
 PR
 XX
 XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
 PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.
 XX
 XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;
 PI
 XX WPI; 2000-147785/13.
 DR
 XX N-PSDB; AAZ88156.
 DR
 XX New male germ cell regulatory factor tesmin expressed in spermatocytes
 PT useful for investigation of germ cell differentiation and sterility -
 PT
 XX Claim 1; Page 50-52; 63pp; Japanese.
 PS
 XX

CC The present sequence represents a male germ cell regulatory factor
 CC expressed specifically in spermatocytes, designated tesmin.
 CC can be used in the investigation of the mechanisms of germ cell
 CC differentiation and sterility.
 XX
 SQ Sequence 299 AA;

Query Match 78.2%; Score 1251; DB 21; Length 299;
 Best Local Similarity 76.3%; Pred. No. 4.2e-111;
 Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

QY 1 MVTCLKGGAGMLCNDCCARELKHLLPOYDDQSFPQSELKPKMTTLVGRLLPVPK 60
 DB 1 mvtclkggagmlcndccarelkahlhlpqyddqsfpqselkpkmttlvgrllvpak 60
 QY 61 LNLITQVDNGALPSAVNGAAPPSPALQGPPTITLGGYDCFSGDFCNCSNLRHEL 120
 DB 61 lnlitqvdngalpsavngaappspalqgpptitlsgydcfsgdfcnscnnlrhel 120
 QY 117 RHELEFKAICLDNRNPEAFQPMKGRGRLGAALKRHSKGCNCRSGCLKNYCEYEAQIMCSS 176
 DB 117 rhelerfkaicldnrnpeafqpmkgrgrrlgaalkrshskgcnckrsgclknycyeyeaqimcss 176
 QY 121 HHDIERFKAICLDNRNPEAFQPMKGRGRLGAALKRHSKGCNCRSGCLKNYCEYEAQIMCSS 180
 DB 121 hhdierfkaicldnrnpeafqpmkgrgrrlgaalkrshskgcnckrsgclknycyeyeaqimcss 180
 QY 177 MCSSICKACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISW 236
 DB 177 mcssickacknyeesperkmlmstphymepgdfesshylsapakfsppklrknrqafscisw 236
 QY 181 MESSICKACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISW 240
 DB 181 messickacknyeesperkmlmstphymepgdfesshylsapakfsppklrknrqafscisw 240
 QY 237 CISWEVVEATCACLQAQEEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
 DB 241 ciswevveatcaccllaqeeaeqhcspslaemileefgrclsqilhieffkskglkme 299

RESULT 3
 ABB62035
 ID ABB62035 standard; Protein; 950 AA.
 XX
 AC ABB62035;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 DE Drosophila melanogaster polypeptide SEQ ID NO 12897.
 XX
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW
 OS Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX
 XX 23-MAR-2001; 2001WO-US09231.
 PF
 XX
 XX 23-MAR-2000; 2000US-191637P.
 PR
 XX 11-JUL-2000; 2000US-0614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR
 XX N-PSDB; ABL06138.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PT
 XX Disclosure; SEQ ID NO 12897; 21pp + Sequence Listing; English.
 XX
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and

DR WPI; 2001-639362/73.
 XX N-PSDB; AAS82145.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 20; SEQ ID No 48317; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 280 AA;

Query Match 22.6%; Score 362; DB 22; Length 280;
 Best Local Similarity 57.9%; Pred. No. 2.8e-26;
 Matches 66; Conservative 15; Mismatches 23; Indels 10; Gaps 2;

Qy 98 YDCFFSGDFCNSCSC---NNLRHELREPKAKACLDNRPEAFQPKMGKGLGAALRH 153
 |||||:||||:| || || || |||||:||||:|: ||
 Db 114 ydcfngfennctnctnynlehenrqaikacldnrpeafkpgkqgesdrrh 173
 |||||:||||:| || || || |||||:||||:|: ||

Qy 154 SKGCNCRSGCLNVCYCPYAKIMSSICKIACKNVEESPERKMLMSTPHYME 207
 |||||:||||:| || || || |||||:||||:|: ||

Db 174 skgcncrsgclnvcycenyliwycr-----yrqilekaiglsgeaql 221
 |||||:||||:| || || || |||||:||||:|: ||

RESULT 6
 ID AAM95330
 AC AAM95330 standard; Protein; 147 AA.
 XX
 AC AAM95330;
 XX
 XX 21-NOV-2001 (first entry)
 DT
 XX
 XX Human reproductive system related antigen SEQ ID NO: 3988.
 DE
 DE Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy.
 KW
 KW Homo sapiens.
 OS
 XX WO200155320-A2.
 XX
 XX 02-AUG-2001.
 PD
 XX 17-JAN-2001; 2001WO-US01339.
 PF
 XX 31-JAN-2000; 2000US-0179065.
 XX 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240950.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
DR N-PSDB; AAL01300.
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
XX Claim 11; SEQ ID NO 3988; 1297pp + Sequence Listing; English.
XX The present invention provides the protein and coding sequences of a

CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention.
XX
SQ Sequence 147 AA;

Query Match 20.1%; Score 322; DB 22; Length 147;
Best Local Similarity 73.8%; Pred. No. 8e-23; Indels 0; Gaps 0;
Matches 62; Conservative 7; Mismatches 15;

Qy 1 MVICOLKGGAQMLCIDNCGARELKALHLLPYQDDSSFPQSELKPKMTLVGRLLPVPK 60
Db 60 mvicqlkggtqmlcidnsrtrelkalhlvpqygdgnylqsdvcpkmtalvgrfipastk 119
Qy 61 LNLITQVDNGALPSAVNGAAPPSG 84
Db 120 lnlitqqlgalpsvngsaifpg 143

RESULT 7
AAG76158
ID AAG76158 standard; Protein; 53 AA.
XX
AC AAG76158;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:6922.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 11.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
XX WPI; 2001-235357/24.
DR N-PSDB; AAG76158.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 8365; 9803pp; English.
XX
XX AAG76158 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAG76158 to AAG77788
XX represent sequences used in the exemplification of the
XX present invention.
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 53 AA;

QY 243 VEATCACLAAQGEAEQHCSPSLAEQMIIEFGRLCSQILHIEFKSKGLKIE 295
DB 1 veatcacciaqgeaekehcsciaeqmleefgrclsqilhtefkskglkme 53

RESULT 8

ABB68888
ID ABB68888 standard; Protein; 243 AA.

XX AC ABB68888;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 33456.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL12991.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Disclosure; SEQ ID NO 33456; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 243 AA;

Query Match 11.6%; Score 186; DB 22; Length 243;

Best Local Similarity 25.4%; Pred. No. 1.5e-09;

Matches 59; Conservative 26; Mismatches 61; Indels 86; Gaps 8;

QY 132 RNPEAFQKMGKRLGAALKRHSGKCRSGGLKYNCEYEAKEIMCSSTCKIACKNYE 191
DB 1 rnpfapfkmgkrlglaalkrhsgkcrsgglkynceyeyakeimcsstckiacnyme 191

Db 7 isvdkadgkkgkqg-gaggyv---k9ccckrsqiknycdyqsmaitkfcrcvgrnte 62
QY 192 -----ESPER-----KMLM 200
DB 63 vrelvdpsnvaknssavkrqkaaaasakaaaaakagidvqgkalqvaastlalpgkalm 122
QY 201 STPHYM-----EPGDFESSHYLSPAKFSGP-----PKLRKN 231
DB 123 ttpkyltvagkp-pmassh-inpipsisrplataatparavkqpaepmpvnliipvrhdd 180
QY 232 RQAFSCISWEVVEATCACLAAQGEAEQHCSPSLAEQMIIEFGRLCSQIL 283
DB 181 rdrnlfvqpvnaallecmliqateadqglnelqvcqlvieefmrgyknail 232

RESULT 9

AAM93348

ID AAM93348 standard; Protein; 251 AA.

XX AC AAM93348;

XX 06-NOV-2001 (first entry)

XX Human polypeptide, SEQ ID NO: 2895.

XX Human; full length cDNA; cDNA synthesis; oligo-capping.

XX Homo sapiens.

XX EF1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2001-524255/58.

XX N-PSDB; AAK94268.

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

XX Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' and 3' ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 251 AA;

Query Match 10.3%; Score 165; DB 22; Length 251;

Best Local Similarity 75.6%; Pred. No. 1.5e-07;

Matches 31; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 MWICQLKGGGAQMLCIDNCGARELKALHLHPQYDDQSSFPQS 41


```

Db 1677 dcassggtltkcsygdncvtvfdsgsvtgrgcse-----svfeesdyc-denp-aycpr 1729
QY 141 MGKGRLLGAALKRHSKGCNCRSGCLKNYCEYEAIKMCSSTC-----KCI 185
Db 1730 c-----nngcntads--qdkyvecivcdsvdncvdsdptqtktrqchercl 1776
QY 186 AC-----KNYEESPERKMLMSTPHYMFGDFESSHYLSPAKFSGPPKLRKNRQAFSCISWE 241
Db 1777 safilplfnetedpsyalirncyddlekdred-----actagskrfcacatcegtknsed 1829
QY 242 VVENTACCLLAQGEA-----EQEHCSPSLAEQMILEEFGRCUSQILH 284
Db 1830 lvasrhscvlcgdcqcdsqgasscnyrehdedyiqfdeersitslg-clselsh 1883

RESULT 12
AA95660
ID AAY95660 standard; Protein; 602 AA.
XX
AC AAY95660;
XX
DT 25-OCT-2000 (first entry)
XX
DE Human Znt2 protein.
XX
KW Znt2; epidermal growth factor-like domain; human;
KW cell differentiation; vulnery; diagnosis; therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..30
FT /label= Signal_peptide
FT /note= "the signal peptide alternatively
FT comprises amino acid residues 9..30"
FT
FT Protein 31..602
FT Domain 31..507
FT /note= "extracellular domain"
FT Domain 508..533
FT /note= "transmembrane domain"
FT Domain 534602
FT /note= "intracellular domain"
FT Region 224..256
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 224..254 or 224..251"
FT Region 272..303
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 272..301 or 272..298"
FT Region 317..351
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 317..349 or 317..346"
FT Region 371..402
FT /note= "laminin-type epidermal growth factor-like
FT domain signature, alternatively residues
FT 317..400 or 317..397"
FT Region 422..449
FT /note= "laminin-type epidermal growth factor-like
FT domain signature"
FT
FT WO200043512-A1.
XX
PD 27-JUL-2000.
XX
PF 20-JAN-2000; 2000WO-US01419.
XX
PR 25-JAN-1999; 99US-0237074.
XX
PA (ZMO ) ZYMOGENETICS INC.
XX

```

```

PI Holloway JL, Lofton-Day CE, Gilbert T;
XX
DR WPI: 2000-491163/43.
DR N-PSDB; AAA50101, AAA50102.
XX
PT Isolated Znt2 nucleic acids and polypeptides which act as epidermal
PT growth factors, useful for the treatment of e.g. kidney and liver
PT disorders, burns, and ulcers and for regulating smooth muscle cell
PT proliferation.
XX
XX Claim 1; Page 89-91; 98pp; English.
XX
CC The present sequence is of that of human Znt2, a new member of the
CC epidermal growth factor-like domain group of proteins. The
CC sequence was deduced from isolated Znt2 cDNA (see AAA50101). In
CC addition to the presence sequence, polypeptides comprising amino
CC acids 31-507 and 31-602 of the sequence are claimed. An expression
CC vector, a recombinant host cell, a method of using the vector to
CC produce Znt2 protein, an antibody or antibody fragment that
CC specifically binds to the polypeptide, and a method of detecting
CC the presence of Znt2 gene products in a biological sample are also
CC all claimed. Znt2 polypeptides may be used to regulate vascular
CC smooth muscle cell proliferation, to restore normal neurological
CC functioning after trauma, to treat ocular disorders, to treat
CC kidney and liver disorders, to promote hair and follicular
CC development, to stimulate growth and differentiation of various
CC epidermal and epithelial cells in vivo and in vitro, for the
CC treatment of burns, ulcers and corneal incisions, and to stimulate
CC wound healing.
XX
SQ Sequence 602 AA;

Query Match 6.1%; Score 97; DB 21; Length 602;
Best Local Similarity 21.8%; Pred. No. 1.5; Indels 74; Gaps 16;
Matches 56; Conservative 26; Mismatches 101;

QY 44 PKPMTTLVGRLLVPVPAKLNILITQVDNGALPSAVNGAAPPSPGALQGPPIKTLSCYCDIFS 103
Db 154 pap-ttpvattvpapt-----tprtptdlpssnsstvlptptateapspppvcncsv 208
QY 104 SGDF-----CN-----SCSCNNLRHELEFKAICACLDNRNPAFQPMKGGRL-----G 147
Db 209 vgslnvnrctqgtgcrcpbgqglh----cetc-----kegfynytsgicqcdcsphg 260
QY 148 AAKL--RHSGKCNCK-----RSGLKNYCEYEAIKMCSICKCIA 186
Db 261 alsipcnssgkckvgvigsicdrqdggygfskngclp--cqcnrnsascdaltg--a 316
QY 187 CKNYEESPERKMLMSTPHYMFGDFESSHYLSPAKFSGPPKLRKNRQAFSCISWEVVEAT 246
Db 317 clncqen-----skgnhce--eckegfygsp-----datkeclrc-pcsavtst 357
QY 247 CACLLAAGE-EAQEHHC 262
Db 358 gscsikssselepecdcq 374

RESULT 13
ABB61083
ID ABB61083 standard; Protein; 1987 AA.
XX
AC ABB61083;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 10041.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX

```

PN W0200171042-A2.
 XX 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 PR (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX N-PSDB; ABL05186.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 10041; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB016176-AB016175), expressed DNA
 CC sequences (AB01840-AB01840) and the encoded proteins
 CC (AB01840-AB01840).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 1987 AA;
 SQ

Query Match 6.1%; Score 97; DB 22; Length 1987;
 Best Local Similarity 20.0%; Pred. No. 6.8;
 Matches 45; Conservative 20; Mismatches 82; Indels 78; Gaps 8;

Qy 65 TOVDNGALPSAV-----NGAAPFSGPALQPPKIPKITSYGYDCDFSGDFGSCNNLR 117
 Db 516 tkveaggtvsslllqlrlnsnvlp--paistesqvl-----cakc---nlq 559
 Qy 118 HELERFKAIAKACLDNRPAFQPKMGKGRGAKLRHSGKCKRSGLKNYCEYEAKIM 177
 Db 560 fdrferfqeln-----khqacsglgsanssqslsqkqerfr 595
 Qy 178 CSSICK-----CIACKNVEESPERKMLMSTPHYWEPGDFESSHYLS 218
 Db 596 csqcsbvqchwhfflmrevhrryiclycnhyvpsveklsl-----hlenkhdldqshfak 651
 Qy 219 PAKFSGPPKLRNROAFSCISWEVVEATCACLQAQGEAEQEHCS 263
 Db 652 dawdqqskedrarhlvcc-----tcqatfvqgsfedhdcs 688

RESULT 14
 AAG09497
 ID AAG09497 standard; Protein; 359 AA.
 XX AAG09497;
 AC AAG09497;
 XX 17-OCT-2000 (first entry)
 DT Arabidopsis thaliana protein fragment SEQ ID NO: 7455.
 DE Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX Arabidopsis thaliana.
 OS

XX PN EP1033405-A2.
 XX PD 06-SEP-2000.
 XX 25-FEB-2000; 2000EP-0301439.
 XX 25-FEB-1999; 99US-0121825.
 XX 05-MAR-1999; 99US-0123180.
 XX 09-MAR-1999; 99US-0123548.
 XX 23-MAR-1999; 99US-0125788.
 XX 25-MAR-1999; 99US-0126264.
 XX 29-MAR-1999; 99US-0126785.
 XX 01-APR-1999; 99US-0127462.
 XX 06-APR-1999; 99US-0128234.
 XX 08-APR-1999; 99US-0128714.
 XX 16-APR-1999; 99US-0129845.
 XX 19-APR-1999; 99US-0130077.
 XX 21-APR-1999; 99US-0130449.
 XX 23-APR-1999; 99US-0130510.
 XX 28-APR-1999; 99US-0130891.
 XX 30-APR-1999; 99US-0131449.
 XX 04-MAY-1999; 99US-0132048.
 XX 05-MAY-1999; 99US-0132407.
 XX 06-MAY-1999; 99US-0132484.
 XX 06-MAY-1999; 99US-0132485.
 XX 06-MAY-1999; 99US-0132486.
 XX 06-MAY-1999; 99US-0132487.
 XX 07-MAY-1999; 99US-0132863.
 XX 11-MAY-1999; 99US-0134256.
 XX 14-MAY-1999; 99US-0134218.
 XX 14-MAY-1999; 99US-0134219.
 XX 14-MAY-1999; 99US-0134221.
 XX 14-MAY-1999; 99US-0134370.
 XX 18-MAY-1999; 99US-0134768.
 XX 20-MAY-1999; 99US-0134941.
 XX 21-MAY-1999; 99US-0135124.
 XX 24-MAY-1999; 99US-0135353.
 XX 25-MAY-1999; 99US-0135629.
 XX 27-MAY-1999; 99US-0136021.
 XX 28-MAY-1999; 99US-0136392.
 XX 01-JUN-1999; 99US-0136782.
 XX 03-JUN-1999; 99US-0137222.
 XX 04-JUN-1999; 99US-0137528.
 XX 07-JUN-1999; 99US-0137502.
 XX 08-JUN-1999; 99US-0137724.
 XX 10-JUN-1999; 99US-0138094.
 XX 10-JUN-1999; 99US-0138540.
 XX 14-JUN-1999; 99US-0138847.
 XX 16-JUN-1999; 99US-0139119.
 XX 16-JUN-1999; 99US-0139452.
 XX 17-JUN-1999; 99US-0139453.
 XX 18-JUN-1999; 99US-0139454.
 XX 18-JUN-1999; 99US-0139455.
 XX 18-JUN-1999; 99US-0139456.
 XX 18-JUN-1999; 99US-0139457.
 XX 18-JUN-1999; 99US-0139458.
 XX 18-JUN-1999; 99US-0139459.
 XX 18-JUN-1999; 99US-0139460.
 XX 18-JUN-1999; 99US-0139461.
 XX 18-JUN-1999; 99US-0139462.
 XX 18-JUN-1999; 99US-0139463.
 XX 18-JUN-1999; 99US-0139730.
 XX 21-JUN-1999; 99US-0139763.
 XX 22-JUN-1999; 99US-0139817.
 XX 23-JUN-1999; 99US-0139899.
 XX 23-JUN-1999; 99US-0140353.
 XX 23-JUN-1999; 99US-0140354.
 XX 24-JUN-1999; 99US-0140695.
 XX 28-JUN-1999; 99US-0140823.
 XX 29-JUN-1999; 99US-0140991.
 XX 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 08-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157553.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 6.0%; Score 96.5; DB 21; Length 359;

Best Local Similarity 32.4%; Pred. No. 0.84; Mismatches 10; Conservative 34; Indels 23; Gaps 7;

QY 131 DRNPEAFOPKMGKRLGAALRHSGK---CNC-KRSGCLKNYCEYEAATMCSSICKCIA 186

Db 238 esnsepfqpetseetl--ckeqskaevccsklcktkckckangsgcgdsqcl 295

QY 187 --CKNYEES--PERKMLMSTPHYMEFGDFESSHYLSPAKFSGP 227

Db 296 skcsnreesakpdkp-----mepidvk-----kpagishdk 327

RESULT 15

AAG09496

ID AAG09496 standard; Protein; 443 AA.

XX AAG09496;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 7454.

DE Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN

XX

PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.

[illegible]

Search completed: July 11, 2002, 08:16:59
Job time: 203 sec

THESE

RESULT 7
 T27827
 hypothetical protein ZK287.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T27827
 R:McMurray, A.
 submitted to the EMBL Data Library, April 1996
 A:Reference number: Z20425
 A:Accession: T27827
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-497 <WIL>
 A:Cross-references: EMBL:Z70757; PIDN:CAA94805.1; GSPDB:GN00023; CESP:ZK287.1
 A:Experimental source: clone ZK287
 C:Genetics:
 A:Gene: CESP:ZK287.1
 A:Map position: 5
 A:Introns: 20/1; 44/2; 99/3; 339/3; 443/2

 Query Match 6.1%; Score 97; DB 2; Length 497;
 Best Local Similarity 26.6%; Pred. No. 1.8;
 Matches 37; Conservative 15; Mismatches 51; Indels 36; Gaps 10;

 QY 82 PSGPALGGPKITLSGYCDCFSGDFCNS--CSCNNLRHELERKAIKACLDNRNPEAFQPK 140
 DB 63 PAKPAKNACAK---SERCDCLRKEGSKLCAKQI-----C--KNKEA--PK 103
 QY 141 MGKRLGAARL-RHSGKCNCK--RSGGLKNYCEYEAKIMCSSICKCIA-CKNYEESPER 196
 DB 104 K-----LAKVAKPTSGCAKGGKQCKVKECACTVYGFCASCKGGDCTNGASK--- 154
 QY 197 KMLMSTPHYMEPGDFESSH 215
 DB 155 ---FSVPKHVQNCFLHKKH 170

 RESULT 8
 MMFFBI
 laminin beta-1 chain precursor - fruit fly (*Drosophila melanogaster*)
 N:Alternate names: laminin chain B1
 C:Species: *Drosophila melanogaster*
 C:date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 19-Jan-2001
 C:Accession: A28783; S14462; B28783
 R:Montell, D.J.; Goodman, C.S.
 Cell 53, 463-473, 1988
 A:title: *Drosophila* substrate adhesion molecule: sequence of laminin B1 chain reveals
 A:Reference number: A28783; MUID:88210471
 A:Accession: A28783
 A:Molecule type: mRNA
 A:Residues: 1-1790 <MONI>
 A:Cross-references: EMBL:M19525
 R:Montell, D.J.; Goodman, C.S.
 submitted to the EMBL Data Library, June 1988
 A:description: *Drosophila* substrate adhesion molecule: sequence of laminin B1 chain r
 A:Reference number: S14462
 A:Molecule type: mRNA
 A:Residues: 1-667, 'L', 669-725, 'VT', 728-947, 950-1790 <MON2>
 A:Cross-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802
 C:Genetics:
 A:Gene: lamB1
 A:Cross-references: FlyBase:FBgn0002527
 A:Map position: 2L 28D
 C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin
 C:Function:
 A:description: Interact with cells and with other basement membrane proteins to promo
 C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
 C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellu
 F:1-26/Domain; signal sequence #status predicted <SIG>

F:27-1790/Product: laminin beta-1 chain #status predicted <MAT>
 F:27-288/Domain: VI <DOM6>
 F:289-561/Domain: V <DOM5>
 F:290-354/Domain: laminin-type EGF-like homology <LE01>
 F:357-417/Domain: laminin-type EGF-like homology <LE02>
 F:420-477/Domain: laminin-type EGF-like homology <LE03>
 F:480-528/Domain: laminin-type EGF-like homology <LE04>
 F:531-561/Domain: laminin-type EGF-like homology <LE05>
 F:562-789/Domain: IV <DOM4>
 F:643-645/Region: cell attachment (R-G-D) motif
 F:790-1189/Domain: III <DOM3>
 F:791-836/Domain: laminin-type EGF-like homology <LE06>
 F:839-882/Domain: laminin-type EGF-like homology <LE07>
 F:885-932/Domain: laminin-type EGF-like homology <LE08>
 F:935-990/Domain: laminin-type EGF-like homology <LE09>
 F:968-972/Region: cell adhesion #status predicted
 F:993-1042/Domain: laminin-type EGF-like homology <LE10>
 F:1045-1093/Domain: laminin-type EGF-like homology <LE11>
 F:1096-1141/Domain: laminin-type EGF-like homology <LE12>
 F:1144-1188/Domain: laminin-type EGF-like homology <LE13>
 F:1190-1407/Domain: II <DOM2>
 F:1408-1434/Domain: alpha <ALP>
 F:1435-1790/Domain: I <DOM1>
 F:51-56/Disulfide bonds: #status predicted
 F:140-203,234,489,593,1053,1248,1303,1332,1343,1475,1495,1517,1583,1646,1705/Binding site
 F:1191,1194,1788/Disulfide bonds: interchain #status predicted

Query Match 6.0%; Score 96.5; DB 1; Length 1790;
 Best Local Similarity 21.0%; Pred. No. 7.2;
 Matches 41; Conservative 15; Mismatches 90; Indels 49; Gaps 6;
 QY 94 TLSGYCDGCGSGDF-----CNSCSG-----NLRHLERLKAIAKALDRNPEAFQPKM 141
 Db 1021 TTGDHCELCRKGDFGDALQNCQCCDFLGNTNTIAHCRFTQCPCCL-----PNV 1072
 QY 142 GKRLGAIAKLRHSK-----GNCRRSGCKLKNYCEYPAKIMSSICKIACK-----N 189
 Db 1073 QGVRCQCAENHAKIAGEGECSCNDPIGALHEQCNSYTGQCCKPFGGRACNQCAH 1132
 QY 190 YEESPERKMLMSTHYMEPGDFESSHYLSPAKFSGLPKLRNQAFAFSCISWEVVEATCAC 249
 Db 1133 YWGNPNKQPCQCDQGAADFQC-----DRETNVCVHEGIGGYKCN 1175
 QY 250 LLAQGEAEQEHCSGP 264
 Db 1176 ECARGYICQFPHCSP 1190
 RESULT 9
 S55554
 male-specific lethal 2 protein - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 21-Jul-2000
 C:Accession: S55554
 R:Zhou, S.; Yang, Y.; Scott, M.J.; Pannuti, A.; Fehr, K.C.; Eisen, A.; Koonin, E.V.; Fod
 EMBO J. 14, 2884-2895, 1995
 A:Title: Male-specific lethal 2, a dosage compensation gene of Drosophila, undergoes sex
 A:Reference number: S55554; MUID:95317307
 A:Accession: S55554
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-769 <ZHO>
 A:Cross-references: EMBL:Z48443; NID:g872110; PIDN:CAA88358.1; PID:g872111
 C:Genetics:
 A:Gene: FlyBase:msl-2
 A:Cross-references: FlyBase:FBgn0005616
 A:Introns: 641/1
 Query Match 6.0%; Score 96; DB 2; Length 769;
 Best Local Similarity 25.7%; Pred. No. 3.3;
 Matches 39; Conservative 17; Mismatches 46; Indels 50; Gaps 9;

QY 121 ERFKAIAKAC-----LDRNPEAFQPKMGKGRGLGAIAKLRHSK-----CNCKRS 163
 Db 471 KRPTLKASQAIAKIEPVSEVKTKVQSGK-GA--LRRIRGKDEKVKPKPKRCRGISG 527
 QY 164 -----CLANYCECYEAKIMSSICKIACKN-----YEES-----PER 196
 Db 528 SSNTLTTCRNSRCPCYKSYNSCAG-CHVCCKPNHKEHYVESEDDDDLEDFEMPKDVPEP 586
 QY 197 KMLMSTHYMEPGDFESSHYLSPAKFSGLPKL 228
 Db 587 MTQSEEPVVAEPQRENS--MAPDSSAPISL 616
 RESULT 10
 T10053
 laminin alpha 5 chain - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000
 C:Accession: T10053
 R:Miner, J.H.; Lewis, R.M.; Sanes, J.R.
 submitted to the EMBL Data Library, November 1997
 A:Reference number: Z16923
 A:Accession: T10053
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-3635 <MIN>
 A:Cross-references: EMBL:U37501; NID:g2599231; PID:g2599232
 C:Genetics:
 A:Gene: Lama5
 C:Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like h
 C:Key words: basement membrane; cell binding; extracellular matrix
 F:1888-1939/Domain: laminin-type EGF-like homology <LEG>
 F:1942-1970/Domain: EGF homology <EGF>

Query Match 6.0%; Score 96; DB 2; Length 3635;
 Best Local Similarity 18.0%; Pred. No. 16;
 Matches 76; Conservative 40; Mismatches 96; Indels 210; Gaps 23;
 QY 14 CIDNCGARELKALHLLPOYDDQSSFSQSELPKPMTWLLVGRLLPVPKALNLTITVDNGLP 73
 Db 117 CLERFGRTLER---ITQDDDDVI-----CTTEYSRIVPL-ENGEIIVSLVNGR-P 161
 QY 74 SAVNGAAPP-----SGPALQCP-----PKITLSGY 98
 Db 162 GALNFSYSPLLRDFTKATNIRLRLTNTLLGLMGLKALRDPVTRRYYSIKDISIGR 221
 QY 99 CDCFSSGDFC-----NSC-----SC 113
 Db 222 CVCHGHADVCDKADPLDPFRLLQACQHQHTCGSCDRCCPGFNQOPWPKPATPDSANEQSC 281
 QY 114 NNLRLH-----ELERFKAIAKALDRNPEAFQPKMGKGRGLGAIAKLRHSKCNCKRS 162
 Db 282 NCHGHAYDCYDPEVDR-----RNASQNDNVYQG--GGVCLDCQHHTTGINGER- 329
 QY 163 GCLKNY-----CECYEAKI---MCSSI-----CK-----CIAC-KNY 190
 Db 330 -CLPGFRAPDQPLDSPHVCRPCDC-ESDFTDGTCDLTGRCYCRRPNTGELCAACAEGY 387
 QY 191 EESPERKMLMSTHYMEPGDFESSHYLSPAKFSGLPKLRNQAFAFSCISWEVVEATCAC 250
 Db 388 TDFPHCYPLPSFPHN-----DTREQVIPA-----QOIVNCDNAA 422
 QY 251 LAQGEAEQEQ-----HC-----SPSLAEQMTLEEFGR 277
 Db 423 GTQGNACRKDPRLGRVCVKPNFRGAHCELCAPGFHGPSCHPCCQSSPGVANSLCDPESGQ 482
 QY 278 CL 279
 Db 483 CM 484

RESULT 11
S61917
protein kinase C (EC 2.7.1.1-) PKCA - Aspergillus niger
C:Species: Aspergillus niger
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S61917
R:Morawetz, R.; Lendenfeld, T.; Mischak, H.; Muehlbauer, M.; Gruber, F.; Goodnight, J. J.; Mol. Gen. Genet. 250, 17-28, 1996
A:Title: Cloning and characterisation of genes (pkc1 and pkcA) encoding protein kinase C
A:Reference number: S61917; MUID:96158841
A:Accession: S61917
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1096 <MOR>
A:Cross-references: EMBL:U10549; NID:g507899; PIDN:AAA97433.1; PID:g507900
C:Genetics:
A:Gene: pkcA
A:Introns: 145/1; 234/1; 428/1; 560/1; 698/1; 874/2; 941/3; 1077/1
C:Superfamily: yeast protein kinase C; protein kinase C zinc-binding repeat homology; protein kinase C; duplication; phospholipid binding; phosphotransferase; serine/threonine kinase
F:460-507/Domain: protein kinase C zinc-binding repeat homology <K21>
F:528-577/Domain: protein kinase C zinc-binding repeat homology <K22>
F:769-1030/Domain: protein kinase homology <KIN>
F:777-785/Region: protein kinase ATP-binding motif

Query Match 5.9%; Score 95; DB 1; Length 1096;
Best Local Similarity 22.6%; Pred. No. 5.8;
Matches 63; Conservative 21; Mismatches 91; Indels 104; Gaps 15;
QY 66 QVNDGALPSAVN-----GAAFTSPGALOGPKKITLSGYCDGFCSSGDFCNCSNNLRHEL 120
Db 359 KMEHGAGPRDTAGGAGSAHPSGPG--GPGAAPSGQYGDGSGAGAPAGNQVMDSW-FAL 415
QY 121 ERFKAI-----KACLDNPAFQPKMGKGRGLGAALR-----HSKG-----C 157
Db 416 EPVGRIYLSMSFAKQLKDRPT-----DIGLNQGAQVRQKEVHEKQGHKFTVQOFYNIM 471
QY 158 NCKRSGLKNYC---ECYEAKTWCSSIC-----KCIACKNYEESPERKMLMSTPHYMEP 208
Db 472 RALCGDFLKYYAGWQACDCKYTKCHKYKVPKVTCKISANTETDP-----517
QY 209 GDFESSHYLSPAKFGPPKLRKNOAFSCISWEVVEATCACLALAOGEAEQEHCSPLAE 268
Db 518 -DEEKINHRIPHFEG-----FSNIS-----ANWCC-----HCG-----545
QY 269 QMILBEFGR-----CLSQLIHTEFKSGKLUKIE 295
Db 546 --YLLPFGKRNKARCTECGLTCHSHCTHLVDPDFCGMSME 582

RESULT 12
T08852
lustrin A - California red abalone
C:Species: Haliotis rufescens (California red abalone)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08852
R:Shen, X.; Belcher, A.M.; Hansma, P.K.; Stucky, G.D.; Morse, D.E. J. Biol. Chem. 272, 32472-32481, 1997
A:Title: Molecular cloning and characterization of lustrin A, a matrix protein from shell
A:Reference number: Z16496; MUID:98070424
A:Accession: T08852
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1428 <SHE>
A:Cross-references: EMBL:AF023459; NID:g2723361; PIDN:AA95154.1; PID:g2723362
A:Experimental source: tissue type mantle (shell and pearl nacre); cell type pallial
C:Superfamily: antileukoproteinas repeat homology
C:Keywords: extracellular matrix; extracellular protein
F:1382-1426/Domain: antileukoproteinas repeat homology <ALP>

Query Match 5.9%; Score 95; DB 2; Length 1428;

Best Local Similarity 22.8%; Pred. No. 7.6;
Matches 56; Conservative 21; Mismatches 81; Indels 88; Gaps 14;
QY 74 SAVNGAAPSFGAL-----OGPKKITLSGYCDGFCSS-----GDFCN-----S 110
Db 409 SCVGAACPRNTACFAAPSGSPAV-----C-CYTSPPRPEPPSPSPPTGDCPGGVN 462
QY 111 CSCNNLR-----HELEFRKAIAKACLDNPAFOP-----KMGKRLGAALKRHSKCNCKRS 162
Db 463 CTAGTCRLVEDCRIGCPAPVKCIDRDPLPPDPVCPVGLG-----DLKLYCGRG 518
QY 163 G--CLKN-YCECYEA-----KIMCS-SICKCIA 186
Db 519 GRRCPWNTYCVIHPADRYAVCCFGSPASAIAPTSAPGVPDPCPGVNVNCTIGVCR 578
QY 187 CANYEESPERKMLMSTPHYMEFGDFESSHYLSPAKFGPPKLRKNOAFSCISWEVVEAT 246
Db 579 NCDYPCPARPTCVD--HSSEP-----SLNCTIGDPALNGKLEEFSCVGGRLCPLN 627
QY 247 CACLLA 252
Db 628 TACLAA 633
RESULT 13
T26998
hypothetical protein Y48B6A.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26998
R:Wall, M.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z20297
A:Accession: T26998
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-716 <WIL>
A:Cross-references: EMBL:AL110490; NID:e1542263; PIDN:CAB54442.1; CESP:Y48B6A.6
A:Gene: CESP:Y48B6A.6
A:Introns: 38/3; 196/3; 437/2; 460/3; 518/1; 617/3; 673/3
Query Match 5.9%; Score 94; DB 2; Length 716;
Best Local Similarity 25.0%; Pred. No. 4.6;
Matches 58; Conservative 33; Mismatches 95; Indels 46; Gaps 10;
QY 22 ELKALHLLPOYDQO-----SSFPQSELPKPMTTLVGRL---LPVPAKLNLIQVD- 68
Db 378 EVKALEAQASDDRPSAPTPIRDSSLPPPPPKPETPLAIRAGPIPTQLEMIHQEDC 437
QY 69 --NGALPSAVNGAAPSFGALOGPKKITLSGYCDGFCSSGDFCNCSNNLRHELERFKAI 126
Db 438 STRSPSPTSVHSGRPSQSPAPVKPSVTVS-----PLGLLCDP-----NL--SIEKPEEM 485
QY 127 KACLDNPAFQPKMGKGRGLGAALKRHSKCNCKRSGLKNCYCEYEAIMCSSICKCIA 186
Db 486 KTEDTKPVETAPAPVDEAEALDADRNRK---INEATCL-----TKIFFSLVSKYFQ 534
QY 187 CANYEESPERKMLMSTPHYMEFGDFESSHYLSPAKFGPPKLRKNOAFSCIS 238
Db 535 NCSFDFLNKRSNGENVPLKMNK---KISLYAEFESEFS-----RKQIQYFSGI 578

RESULT 14

A71928

cag island protein - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: A71928

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.

C;superfamily: laminin alpha 4 chain, laminin 9 repeat homolog, laminin type 9
C;keywords: basement membrane; cell binding; coiled coil; extracellular matrix; glyco
F;1-24/Domain: signal sequence #status predicted <SIG>

F; 132-137/Domain: laminin-type EGF-like homology <LE2>
F; 132-184/Domain: laminin-type EGF-like homology <LE2>
F; 187-238/Domain: laminin-type EGF-like homology <LE3>
F; 241-265/Domain: laminin-type EGF-like homology #status atypical <LE4>

F; 862-1031/Domain: laminin G repeat homology <LG1>
F; 1068-1223/Domain: laminin G repeat homology <LG2>
F; 1352-1398/Domain: laminin G repeat homology <LG3>

F:1866 -1816/Domain: laminin G repeat homology <L6>
F:1866 -1816/Domain: laminin G repeat homology <L6>
F:1404,215,308,458,524,550,571,574,631,639,735,
F:106,269/Disulfide bonds: interchain #status predicted
F:266,269/Disulfide bonds: interchain #status predicted

Query Match 5.8%; Score 93; DB 1; Length 1816;
Best Local Similarity 23.8%; Pred. NO. 14;

QY	2	VICQLKGGAQMLCIDNCG	----	ARELKALH	----	LLPOY	----	DDQSFPQSELKPKMTTLV	51
			:		:				

QY 52 GRLLPVPAKNLITQV--DNGALSAVN-GAAPFS-----GPAQQPPKLTLSGYDCFFS 104
 | | ||| | | | |
Db 134 ---CPLHLANFAESCYSRKNGAVRGICNENYAGNCERCAPGYGNPLLI-----180

```
QY      105 GDFCNCSN-----NLRHE--LRFKATKACIDRNPAFQ-PMKGKGRGAACL-RHSK 155
|       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
D6      181 GSTCKKKDCSGNSDPNLIFEDCDEVTGQCRL-RTTTFGKGCERCAFYGADRIAKNC 239
```

QY 156 GCNCKRSGCLNYCECYEAKIMCSSIC-KCI-----ACKNYBESPERKMLMST--P 203
 ||| | | | |
Db 240 VCNCGGPGCDSVTEGLEEGFEPTGTCDRCVWDLTDDLRLAALSIEEGKSGLVSLSVGAA 299
 ||| | | | |

QY 204 HYMEPGDFESSHYLSPAKFGSPPKLRKNQAFSCISWEVVEATCAllAQEEAEQEHCS 263

Db 300 AHRHVNEINATILYLLTKLSE----RENYALRRKIQINNAENTMKSLSDYEEELVKEKQ 355

Qy	264	PSLAEQMILEE	274
		:::	
Db	356	ASRKGLVQKE	366

Search completed: July 11, 2002, 08:15:20

Search completed: July 11, 2002, 08:15:20
Job time: 104 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:15:26 ; Search time 30.23 seconds
(without alignments)
377.845 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRLSQILHIEFKSKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1599	100.0	295	1 MTL5_MOUSE	Q9wtj6 mus musculu
2	1251	78.2	299	1 MTL5_HUMAN	Q9y415 homo sapien
3	96.5	6.0	1790	1 LMB1_DROME	P11046 drosophila
4	96	6.0	773	1 MSL2_DROME	P50534 drosophila
5	96	6.0	3718	1 LMA5_MOUSE	Q61001 mus musculu
6	95	5.9	1096	1 KPC1_ASPNG	Q00078 aspergillus
7	93	5.8	1816	1 LMA4_HUMAN	Q16363 homo sapien
8	92	5.8	798	1 ITB1_XENLA	P12606 xenopus lae
9	91	5.7	493	1 OC90_HUMAN	Q02509 homo sapien
10	91	5.7	747	1 E2H1_HUMAN	Q92800 homo sapien
11	91	5.7	2907	1 FBN2_MOUSE	Q61555 mus musculu
12	89	5.6	1786	1 LMB1_MOUSE	P02469 mus musculu
13	88	5.5	798	1 ITB0_XENLA	P12607 xenopus lae
14	88	5.5	1786	1 LMB1_HUMAN	P07942 homo sapien
15	87	5.4	290	1 Y070_NPVAC	P41470 autographa
16	87	5.4	501	1 TRA2_HUMAN	P12933 homo sapien
17	87	5.4	652	1 SVG_SCHPO	Q10179 schizosach
18	86.5	5.4	611	1 LEM2_CANFA	P33730 canis famil
19	86	5.4	610	1 LEM2_HUMAN	P16581 homo sapien
20	86	5.4	747	1 E2H1_MOUSE	P70351 mus musculu
21	86	5.4	830	1 LEM3_HUMAN	P16109 homo sapien
22	85.5	5.3	769	1 ITB7_MOUSE	P05107 homo sapien
23	85.5	5.3	806	1 ITB7_MOUSE	P26011 mus musculu
24	85	5.3	398	1 MUB1_XENLA	P38565 xenopus lae
25	85	5.3	746	1 E2H2_MOUSE	Q61188 mus musculu
26	85	5.3	787	1 ITB6_MOUSE	Q920t9 mus musculu
27	84.5	5.3	169	1 KRG3_HUMAN	P26371 homo sapien
28	84.5	5.3	720	1 NRG3_HUMAN	P56975 homo sapien
29	84.5	5.3	1639	1 LMCL_DROME	P15215 drosophila
30	84.5	5.3	1816	1 LMA4_MOUSE	P97927 mus musculu
31	84	5.3	769	1 ITB2_BOVIN	P32592 bos taurus
32	84	5.3	2531	1 NTC1_MOUSE	Q01705 mus musculu
33	84	5.3	2531	1 NTC1_RAT	Q07008 rattus norv

34	84	5.3	3707	1 PGBM_MOUSE	Q05793 mus musculu
35	84	5.3	3828	1 TRX_DROVI	Q24742 drosophila
36	83.5	5.2	419	1 VEGC_HUMAN	P49767 homo sapien
37	83.5	5.2	769	1 ITB2_PIG	P53714 sus scrofa
38	83.5	5.2	798	1 ITB7_HUMAN	P26010 homo sapien
39	83.5	5.2	2766	1 THYG_MOUSE	O08710 mus musculu
40	83	5.2	423	1 MKR2_SEROU	Q9dd48 seriola qui
41	83	5.2	746	1 E2H2_HUMAN	Q15910 homo sapien
42	83	5.2	771	1 ITB2_MOUSE	P11835 mus musculu
43	83	5.2	1877	1 PKK5_MOUSE	Q04592 mus musculu
44	82.5	5.2	445	1 AD18_RAT	P97776 rattus norv
45	82.5	5.2	768	1 LEM3_MOUSE	Q01102 mus musculu

ALIGNMENTS

```
RESULT 1
MTL5_MOUSE
ID MTL5_MOUSE STANDARD; PRT; 295 AA.
AC Q9WTJ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Tesmin (Metallothionein-like 5, testis-specific) (Testis-specific
DE metallothionein-like protein).
GN MTL5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, tesmin, is an
RT early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -!- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES
CC WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED
CC PROGRESSIVELY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U77383; AAD24667.1; -.
DR EMBL; U61716; AAD24666.1; -.
DR MGD; MGI:1340029; Mtl5.
KW Spermatogenesis.
FT DOMAIN 99. 187 CYS-RICH.
SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;
```

```
Query Match 100.0%; Score 1599; DB 1; Length 295;
Best Local Similarity 100.0%; Pred. No. 8.2e-132;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVICQLKGAQMLCIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTTLVGRLLPVPAP 60
Db 1 MVICQLKGAQMLCIDNCGARELKALHLLPQYDDQSSFFQSELPKPMPTTLVGRLLPVPAP 60
QY 61 LNLITQVDNGALPSAVNGAAPPSPGALQGPKKITLTSYGCDCPSSGDFGFCSCSCNNLRHEL 120
Db 61 LNLITQVDNGALPSAVNGAAPPSPGALQGPKKITLTSYGCDCPSSGDFGFCSCSCNNLRHEL 120
```

QY 121 ERKATKACLDNRNPEAFQPKMGKGRGAAKLRHSGKCGNCRGCLKNYCECYEAKIMCSS 180
DB 121 ERKATKACLDNRNPEAFQPKMGKGRGAAKLRHSGKCGNCRGCLKNYCECYEAKIMCSS 180

QY 181 IKCIACKNYEESPERKMLMSTPHYMEPGDFESSHVLSPAKTSGPPKLRNROAFSCISW 240
DB 181 IKCIACKNYEESPERKMLMSTPHYMEPGDFESSHVLSPAKTSGPPKLRNROAFSCISW 240

QY 241 EVVEATCACLGAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
DB 241 EVVEATCACLGAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295

RESULT 2
MTLS5_HUMAN
ID MTLS5_HUMAN STANDARD; PRT; 299 AA.
AC Q9Y415;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Testin (Metallothionein-like 5, testis-specific) (Testis-specific metallothionein-like protein).
GN MTLS5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99208669; PubMed=10191092;
RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
RT "A novel testis-specific metallothionein-like protein, testin, is an early marker of male germ cell differentiation.";
RL Genomics 57:130-136(1999).
CC -1- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U86074; AAD24668.1; -
DR MIM: 604374; -
KW Spermatogenesis.
FT DOMAIN 99 191 CYS-RICH.
SQ SEQUENCE 299 AA; 33110 MW; 8F679E140F969C11 CRC64;

Query Match 78.2%; Score 1251; DB 1; Length 299;
Best Local Similarity 76.3%; Pred. No. 1.4e-101;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

QY 1 MVICOLGGAGMCIIDNCAGARELKALHLLPOYDQSQFPOSELPKMTTLVGRLLPVPK 60
DB 1 MVICOLGGAGMCIIDNCAGARELKALHLLPOYDQSQFPOSELPKMTTLVGRLLPVPK 60

QY 61 LNLITQVNDNGALPSAVNGAAPPSPGALOGPPKIFLGSYDCDFSSGDFCNSCS-----CNL 116
DB 61 LNLITQVNDNGALPSAVNGAAPPSPGALOGPPKIFLGSYDCDFSSGDFCNSCS-----CNL 120

QY 117 RHLEERKAIKACLDNRNPEAFQPKMGKGRGAAKLRHSGKCGNCRGCLKNYCECYEAKI 176
DB 117 RHLEERKAIKACLDNRNPEAFQPKMGKGRGAAKLRHSGKCGNCRGCLKNYCECYEAKI 180

QY 177 MCSSTICKIACKNYEESPERKMLMSTPHYMEPGDFESSHVLSPAKTSGPPKLRNROAFS 236
DB 177 MCSSTICKIACKNYEESPERKMLMSTPHYMEPGDFESSHVLSPAKTSGPPKLRNROAFS 240

QY 237 CISWEVVEATCACLGAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 295
DB 241 CISWEVVEATCACLGAQGEAEQHCSPSLAEQMIIEFGRCLSQILHIEFKSKGLKIE 299

RESULT 3
LMBL_DROME
ID LMBL_DROME STANDARD; PRT; 1790 AA.
AC P11046; Q26328; Q9XZT4; Q9VLM6;
DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laminin beta-1 chain precursor (Laminin B1 chain).
GN LANB1 OR LAMB1 OR CG7123.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=94000382; PubMed=8397815;
RA Gow C.-H., Chang H.-Y., Lih C.-J., Chang T.-W., Hui C.-F.;
RT "Analysis of the Drosophila gene for the laminin B1 chain.";
RL DNA Cell Biol. 12:573-587(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88210471; PubMed=3365769;
RA Montell D.J., Goodman C.S.;
RT "Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals domains of homology with mouse.";
RL Cell 53:463-473(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Cocayne J.D., Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Mortman J.R., Vandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., April J.F., Agayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D., Balles R.M., Basu A., Baxendale J., Berman B.P., Bhandari D., Bolshakov S., Beeson K.Y., Benos P.V., Bouton J., Brokstein P., Brothier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M., Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ

CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF

CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING

CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.

CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE

CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND

CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE

CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.

CC -1- SUBCELLULAR LOCATION: Extracellular.

CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR

CC COMPONENT).

CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT

CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.

CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M95811; AAD19752.1; -

DR EMBL; M19525; AAA28663.1; -

DR EMBL; AE003618; AAF52563.1; -

DR PIR; A28783; MMRFB1.

DR HSP; P02468; TITLE

DR FlyBase; FBgn0002527; LanB1.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001886; LamNT.

DR InterPro; IPR002049; Laminin_EGF.

DR Pfam; PF00053; laminin_EGF; 13.

DR Pfam; PF00055; laminin_Nterm; 1.

DR PRINTS; PR00011; EGF_LAMININ.

DR ProDom; PD002082; LamNT; 1.

DR SMART; SM00180; EGF_Lam; 11.

DR SMART; SM00001; EGF_Like; 1.

DR SMART; SM00136; LamNT; 1.

DR PROSITE; PS00022; EGF_1; 10.

DR PROSITE; PS01186; EGF_2; 2.

DR PROSITE; PS01248; LAMININ_TYPE_EGF; 12.

KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;

KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

FT SIGNAL 1 25

FT CHAIN 26 1790 LAMININ BETA-1 CHAIN.

FT DOMAIN 43 289 LAMININ N-TERMINAL (DOMAIN VI).

FT DOMAIN 290 356 LAMININ EGF-LIKE 1.

FT DOMAIN 357 419 LAMININ EGF-LIKE 2.

FT DOMAIN 420 479 LAMININ EGF-LIKE 3.

FT DOMAIN 480 530 LAMININ EGF-LIKE 4.

FT DOMAIN 531 561 LAMININ EGF-LIKE 5 (INCOMPLETE).

FT DOMAIN 562 789 LAMININ DOMAIN IV.

FT DOMAIN 791 838 LAMININ EGF-LIKE 6.

FT DOMAIN 839 884 LAMININ EGF-LIKE 7.

FT DOMAIN 885 934 LAMININ EGF-LIKE 8.

FT DOMAIN 935 992 LAMININ EGF-LIKE 9.

FT DOMAIN 993 1044 LAMININ EGF-LIKE 10.

FT DOMAIN 1045 1095 LAMININ EGF-LIKE 11.

FT DOMAIN 1096 1143 LAMININ EGF-LIKE 12.

FT DOMAIN 1144 1190 LAMININ EGF-LIKE 13.

FT DOMAIN 1191 1407 DOMAIN II.

FT DOMAIN 1408 1434 DOMAIN ALPHA.

FT DOMAIN 1435 1790 DOMAIN I.

FT DOMAIN 1257 1407 COILED COIL (POTENTIAL).

FT DOMAIN 1455 1507 COILED COIL (POTENTIAL).

FT DOMAIN 1542 1563 COILED COIL (POTENTIAL).

FT DOMAIN 1610 1764 COILED COIL (POTENTIAL).

FT DISULFID 290 292 BY SIMILARITY.

FT DISULFID 322 320 BY SIMILARITY.

FT DISULFID 334 331 BY SIMILARITY.

FT DISULFID 357 354 BY SIMILARITY.

FT DISULFID 387 384 BY SIMILARITY.

FT DISULFID 396 396 BY SIMILARITY.

FT DISULFID 417 417 BY SIMILARITY.

FT DISULFID 420 433 BY SIMILARITY.

FT DISULFID 422 448 BY SIMILARITY.

FT DISULFID 450 459 BY SIMILARITY.

FT DISULFID 462 477 BY SIMILARITY.

FT DISULFID 480 493 BY SIMILARITY.

FT DISULFID 482 500 BY SIMILARITY.

FT DISULFID 502 511 BY SIMILARITY.

FT DISULFID 514 528 BY SIMILARITY.

FT DISULFID 791 803 BY SIMILARITY.

FT DISULFID 793 810 BY SIMILARITY.

FT DISULFID 812 821 BY SIMILARITY.

FT DISULFID 824 836 BY SIMILARITY.

FT DISULFID 839 851 BY SIMILARITY.

FT DISULFID 841 858 BY SIMILARITY.

FT DISULFID 860 869 BY SIMILARITY.

FT DISULFID 872 882 BY SIMILARITY.

FT DISULFID 885 894 BY SIMILARITY.

FT DISULFID 887 901 BY SIMILARITY.

FT DISULFID 904 913 BY SIMILARITY.

FT DISULFID 916 932 BY SIMILARITY.

FT DISULFID 935 951 BY SIMILARITY.

FT DISULFID 937 962 BY SIMILARITY.

FT DISULFID 964 973 BY SIMILARITY.

FT DISULFID 976 990 BY SIMILARITY.

FT DISULFID 993 1007 BY SIMILARITY.

FT DISULFID 1014 1014 BY SIMILARITY.

FT DISULFID 1017 1026 BY SIMILARITY.

FT DISULFID 1029 1042 BY SIMILARITY.

FT DISULFID 1045 1059 BY SIMILARITY.

FT DISULFID 1047 1066 BY SIMILARITY.

FT DISULFID 1068 1077 BY SIMILARITY.

FT DISULFID 1080 1093 BY SIMILARITY.

FT DISULFID 1096 1108 BY SIMILARITY.

FT DISULFID 1098 1115 BY SIMILARITY.

FT DISULFID 1117 1126 BY SIMILARITY.

FT DISULFID 1129 1141 BY SIMILARITY.

FT DISULFID 1144 1156 BY SIMILARITY.

FT DISULFID 1146 1163 BY SIMILARITY.

FT DISULFID 1165 1174 BY SIMILARITY.

FT DISULFID 1177 1188 BY SIMILARITY.

FT DISULFID 1191 1191 INTERCHAIN (PROBABLE).

FT DISULFID 1194 1194 INTERCHAIN (PROBABLE).

FT DISULFID 1788 1788 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 643 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 203 203 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 489 489 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 593 593 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1248 1248 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 6.0%; Score 96.5; DB 1; Length 1790;

Best Local Similarity 21.0%; Pred. No. 1.9;

Matches 41; Conservative 15; Mismatches 90; Indels 49; Gaps 6;

QY 94 TSLGYDCDFSSGDF-----CNSCS-----NLRHELEKAIKACLDNRPEARQPKM 141

DB 1021 TTGDHCEICKOGFFDALQONCQCCEDFLGTNTTAHCDRTFGQCPCL-----PNV 1072

QY 142 GKGRIGAARLHRSK-----GCNCKRSGLKNYCEYEAKIMCSSICKIACK-----N 189

DB 1073 QGVRCDCAENHWKIASGEGCSCNCDPIGALHEOCNSYTGOCCKPGFGGRACNOCQAH 1132

QY 190 YEESPERKMLMSTHYMEPCGDFESSHYLSPAKFSGPPKLRNKAQFSCISWEVVEATCAC 249
 DB 1133 YWGNPNKQCPCQDQFGAADFQC-----DRETCNCVCEGGYKCN 1175
 QY 250 LLAQGEAEQEHCS 264
 DB 1176 ECARGYIGQFPHCSP 1190

RESULT 4
 MSL2_DROME STANDARD; PRT; 773 AA.
 AC P50534;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Male-specific lethal-2 protein.
 GN MSL-2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95300219; PubMed=7781064;
 RA Kelley R.L., Solovyeva I., Lyman L.M., Richman R., Solovyev V.,
 RA Kuroda M.I.;
 RT "Expression of msl-2 causes assembly of dosage compensation
 RT regulators on the X chromosomes and female lethality in Drosophila."
 RL Cell 81:867-877(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95317307; PubMed=7796814;
 RA Zhou S., Yang Y., Scott M.J., Pannuti A., Fehr K.C., Eisen A.,
 RA Koonin E.V., Fouts D.L., Wrightman R., Manning J.E., Lucchesi J.C.;
 RT "Male-specific lethal 2, a dosage compensation gene of Drosophila,
 RT undergoes sex-specific regulation and encodes a protein with a RING
 RT finger and a metallothionein-like cysteine cluster."
 RL EMBO J. 14:2884-2895(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96017637; PubMed=7588059;
 RA Beshaw G.J., Baker B.S.;
 RT "The msl-2 dosage compensation gene of Drosophila encodes a putative
 RT DNA-binding protein whose expression is sex specifically regulated
 RT by Sex-lethal."
 RL Development 121:3245-3258(1995).
 CC -1- FUNCTION: THE MSL PROTEINS ARE ESSENTIAL FOR ELEVATING
 CC TRANSCRIPTION OF THE SINGLE X CHROMOSOME IN THE MALE (X CHROMOSOME
 CC DOSAGE COMPENSATION). MSL-2 IS REQUIRED FOR TRANSLATION AND/OR
 CC STABILITY OF MSL-1 IN MALES. BOTH MSL-1 AND MSL-2 BIND TO HUNDREDS
 CC OF SITES ALONG THE MALE X CHROMOSOME, BUT NOT THE FEMALE X OR THE
 CC AUTOSOMES IN EITHER SEX.
 CC -1- SUBUNIT: MSL-2 SEEMS TO FORM A TIGHT COMPLEX WITH MSL-1.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- DEVELOPMENTAL STAGE: THE MSL-2 PROTEIN IS ONLY PRODUCED IN MALES.
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L42553; AAA75573.1; -
 CC EMBL; Z4843; CAA8358.1; -
 CC EMBL; X89241; CAA61529.1; -
 CC FlyBase; FBgn0005616; msl-2.
 CC InterPro; IPR001841; Znf_fing.

DR SMART; SM00184; RING; 1.
 DR PROSITE; PS00518; ZF_RING_1; 1.
 DR PROSITE; PS00089; ZF_RING_2; 1.
 KW Zinc-finger; Nuclear protein; DNA-binding; Coiled coil.
 FT ZN_FING 41 85 RING-TYPE
 FT DOMAIN 283 519 COILED COIL (POTENTIAL).
 FT CONFLICT 280 280 T -> N (IN REF. 2).
 FT CONFLICT 370 373 MISSING (IN REF. 2).
 FT CONFLICT 371 373 LKT -> VEE (IN REF. 3).
 FT CONFLICT 380 380 P -> Q (IN REF. 2).
 FT CONFLICT 403 403 T -> I (IN REF. 2).
 FT CONFLICT 403 403 T -> A (IN REF. 2).
 FT CONFLICT 405 405 T -> A (IN REF. 2).
 FT CONFLICT 423 423 H -> Q (IN REF. 3).
 FT CONFLICT 429 429 Q -> H (IN REF. 2).
 FT CONFLICT 431 431 A -> E (IN REF. 2 AND 3).
 FT CONFLICT 433 433 S -> P (IN REF. 2 AND 3).
 FT CONFLICT 439 439 E -> V (IN REF. 2).
 FT CONFLICT 443 443 A -> P (IN REF. 2).
 FT CONFLICT 471 471 T -> A (IN REF. 3).
 FT CONFLICT 591 591 M -> L (IN REF. 3).
 FT CONFLICT 642 642 E -> D (IN REF. 2).
 SQ SEQUENCE 773 AA; 84841 MW; AE38BA76A4174225 CRC64;

Query Match 6.0%; Score 96; DB 1; Length 773;
 Best Local Similarity 25.7%; Pred. No. 0.81;
 Matches 39; Conservative 17; Mismatches 46; Indels 50; Gaps 9;

QY 121 ERFKAIKAC---LDRNPEAFOPKMGKGLGAALRHSGK-----CNCKRSG 163
 DB 475 KRTRLKASAAKLEPVEVTKVQSGK-GA--LRRIRGDKKEKVKPKPKRCGISG 531
 QY 164 -----CLKNYCEYBAKIMCSSICKIACKN-----YEES-----PER 196
 DB 532 SSNTLTTCRNSRCPCKYSYNSCAG-CHVCCKNPKHEDYVEDEDDLEDFEMPKDYPEP 590
 QY 197 KMLMSTHYMEPCGDFESSHYLSPAKFSGPPKL 228
 DB 591 MTQSEEPVVAEPROEENS--MAPPDSSAPISL 620

RESULT 5
 LMA5_MOUSE STANDARD; PRT; 3718 AA.
 AC Q61001; Q9JHQ6;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Laminin alpha-5 chain precursor.
 GN LAMA5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE OF 1-92 FROM N.A.
 RA Timpl R., Sasaki T.;
 RT "Completion of the N-terminal sequence of the murine Laminin alpha 5
 RT chain."
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 84-3718 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
 RX MEDLINE=96081906; PubMed=7499364;
 RA Miner J.H., Lewis R.M., Sanes J.R.;
 RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
 RT expression in adult mouse tissues."
 RL J. Biol. Chem. 270:28523-28526(1995).
 RN [3]
 RP REVISIONS.
 RA Miner J.H., Lewis R.M., Sanes J.R.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ

IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.

- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.

- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.

- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT MEMBRANES (MAJOR COMPONENT).

- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.

- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.

- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.

- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.

- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AJ293593; CAB99255.1; -
 EMBL: U37501; AAC53430.1; -
 HSSP: P02468; 1TLE
 MGD: MGI:105382; Lama5
 InterPro: IPR000561; EGF-like.
 InterPro: IPR001886; LamNT.
 InterPro: IPR000034; Laminin_B.
 InterPro: IPR002049; Laminin_EGF.
 InterPro: IPR001791; Laminin_G.
 Pfam: PF00052; laminin_B; 1.
 Pfam: PF00053; laminin_EGF; 19.
 Pfam: PF00054; laminin_G; 2.
 Pfam: PF00055; laminin_Nterm; 1.
 PRINTS: PR00011; EGF_LAMININ.
 ProDom: PD002082; LamNT; 1.
 ProDom: PD003031; Laminin_B; 1.
 SMART: SM00180; EGF_Lam; 17.
 SMART: SM00001; EGF_like; 2.
 SMART: SM00281; LamB; 1.
 SMART: SM00282; LamG; 5.
 SMART: SM00136; LamNT; 1.
 PROSITE: PS00022; EGF_1; 19.
 PROSITE: PS01186; EGF_2; 3.
 PROSITE: PS01248; LAMININ_TYPE_EGF; 19.
 PROSITE: PS50025; LAM_G_DOMAIN; 5.
 Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal.

SIGNAL 1 40
 POTENTIAL.

FT CHAIN 41 3718 LAMININ ALPHA-5 CHAIN.
 FT DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 305 363 LAMININ EGF-LIKE 1.
 FT DOMAIN 364 433 LAMININ EGF-LIKE 2.
 FT DOMAIN 434 479 LAMININ EGF-LIKE 3.
 FT DOMAIN 500 546 LAMININ EGF-LIKE 4.
 FT DOMAIN 547 592 LAMININ EGF-LIKE 5.
 FT DOMAIN 593 637 LAMININ EGF-LIKE 6.
 FT DOMAIN 638 682 LAMININ EGF-LIKE 7.
 FT DOMAIN 683 728 LAMININ EGF-LIKE 8.
 FT DOMAIN 729 781 LAMININ EGF-LIKE 9.
 FT DOMAIN 782 833 LAMININ EGF-LIKE 10.
 FT DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).
 FT DOMAIN 856 1442 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 1443 1488 LAMININ EGF-LIKE 12.

1489 DOMAIN FT DOMAIN 1489 1532 LAMININ EGF-LIKE 13.
 1533 DOMAIN FT DOMAIN 1533 1581 LAMININ EGF-LIKE 14.
 1582 DOMAIN FT DOMAIN 1582 1632 LAMININ EGF-LIKE 15.
 1633 DOMAIN FT DOMAIN 1633 1642 LAMININ EGF-LIKE 16 (N-TERMINAL).
 1643 DOMAIN FT DOMAIN 1643 1831 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 1832 DOMAIN FT DOMAIN 1832 1864 LAMININ EGF-LIKE 16 (C-TERMINAL).
 1865 DOMAIN FT DOMAIN 1865 1914 LAMININ EGF-LIKE 17.
 1915 DOMAIN FT DOMAIN 1915 1970 LAMININ EGF-LIKE 18.
 1971 DOMAIN FT DOMAIN 1971 2024 LAMININ EGF-LIKE 19.
 2025 DOMAIN FT DOMAIN 2025 2071 LAMININ EGF-LIKE 20.
 2072 DOMAIN FT DOMAIN 2072 2118 LAMININ EGF-LIKE 21.
 2119 DOMAIN FT DOMAIN 2119 2168 LAMININ EGF-LIKE 22.
 2169 DOMAIN FT DOMAIN 2169 2735 DOMAIN II AND I.
 2736 DOMAIN FT DOMAIN 2736 2933 LAMININ G-LIKE 1.
 2934 DOMAIN FT DOMAIN 2934 3119 LAMININ G-LIKE 2.
 3120 DOMAIN FT DOMAIN 3120 3296 LAMININ G-LIKE 3.
 3297 DOMAIN FT DOMAIN 3297 3511 LAMININ G-LIKE 4.
 3512 DOMAIN FT DOMAIN 3512 3689 LAMININ G-LIKE 5.
 3690 DOMAIN FT DOMAIN 3690 2257 COILED COIL (POTENTIAL).
 2258 DOMAIN FT DOMAIN 2258 2464 COILED COIL (POTENTIAL).
 2465 DOMAIN FT DOMAIN 2465 2621 COILED COIL (POTENTIAL).
 2622 DOMAIN FT DOMAIN 2622 2705 COILED COIL (POTENTIAL).
 2706 DOMAIN FT DOMAIN 2706 1725 CELL ATTACHMENT SITE (POTENTIAL).
 1727 DOMAIN FT SITE 1727 1841 CELL ATTACHMENT SITE (POTENTIAL).
 1842 DOMAIN FT SITE 1842 305 BY SIMILARITY.
 306 DOMAIN FT DISULFID 306 327 BY SIMILARITY.
 328 DOMAIN FT DISULFID 328 338 BY SIMILARITY.
 339 DOMAIN FT DISULFID 339 341 BY SIMILARITY.
 342 DOMAIN FT DISULFID 342 364 BY SIMILARITY.
 365 DOMAIN FT DISULFID 365 373 BY SIMILARITY.
 374 DOMAIN FT DISULFID 374 398 BY SIMILARITY.
 399 DOMAIN FT DISULFID 399 410 BY SIMILARITY.
 411 DOMAIN FT DISULFID 411 431 BY SIMILARITY.
 432 DOMAIN FT DISULFID 432 445 BY SIMILARITY.
 446 DOMAIN FT DISULFID 446 452 BY SIMILARITY.
 453 DOMAIN FT DISULFID 453 463 BY SIMILARITY.
 464 DOMAIN FT DISULFID 464 476 BY SIMILARITY.
 477 DOMAIN FT DISULFID 477 512 BY SIMILARITY.
 513 DOMAIN FT DISULFID 513 521 BY SIMILARITY.
 522 DOMAIN FT DISULFID 522 532 BY SIMILARITY.
 533 DOMAIN FT DISULFID 533 544 BY SIMILARITY.
 545 DOMAIN FT DISULFID 545 559 BY SIMILARITY.
 560 DOMAIN FT DISULFID 560 566 BY SIMILARITY.
 567 DOMAIN FT DISULFID 567 577 BY SIMILARITY.
 578 DOMAIN FT DISULFID 578 590 BY SIMILARITY.
 591 DOMAIN FT DISULFID 591 605 BY SIMILARITY.
 606 DOMAIN FT DISULFID 606 611 BY SIMILARITY.
 612 DOMAIN FT DISULFID 612 622 BY SIMILARITY.
 623 DOMAIN FT DISULFID 623 635 BY SIMILARITY.
 636 DOMAIN FT DISULFID 636 650 BY SIMILARITY.
 651 DOMAIN FT DISULFID 651 667 BY SIMILARITY.
 668 DOMAIN FT DISULFID 668 680 BY SIMILARITY.
 681 DOMAIN FT DISULFID 681 695 BY SIMILARITY.
 696 DOMAIN FT DISULFID 696 702 BY SIMILARITY.
 703 DOMAIN FT DISULFID 703 713 BY SIMILARITY.
 714 DOMAIN FT DISULFID 714 726 BY SIMILARITY.
 727 DOMAIN FT DISULFID 727 1443 BY SIMILARITY.
 1444 DOMAIN FT DISULFID 1444 1455 BY SIMILARITY.
 1456 DOMAIN FT DISULFID 1456 1462 BY SIMILARITY.
 1463 DOMAIN FT DISULFID 1463 1473 BY SIMILARITY.
 1474 DOMAIN FT DISULFID 1474 1486 BY SIMILARITY.
 1487 DOMAIN FT DISULFID 1487 1533 BY SIMILARITY.
 1534 DOMAIN FT DISULFID 1534 1555 BY SIMILARITY.
 1556 DOMAIN FT DISULFID 1556 1566 BY SIMILARITY.
 1567 DOMAIN FT DISULFID 1567 1579 BY SIMILARITY.
 1580 DOMAIN FT DISULFID 1580 1594 BY SIMILARITY.
 1595 DOMAIN FT DISULFID 1595 1601 BY SIMILARITY.
 1602 DOMAIN FT DISULFID 1602 1612 BY SIMILARITY.
 1613 DOMAIN FT DISULFID 1613 1630 BY SIMILARITY.
 1631 DOMAIN FT DISULFID 1631 1655 BY SIMILARITY.
 1656 DOMAIN FT DISULFID 1656 1874 BY SIMILARITY.
 1875 DOMAIN FT DISULFID 1875 1881 BY SIMILARITY.
 1882 DOMAIN FT DISULFID 1882 1893 BY SIMILARITY.
 1894 DOMAIN FT DISULFID 1894 1912 BY SIMILARITY.
 1913 DOMAIN FT DISULFID 1913 1915 BY SIMILARITY.
 1916 DOMAIN FT DISULFID 1916 1930 BY SIMILARITY.

FT DISULFID 1917 1939 BY SIMILARITY.
 FT DISULFID 1941 1950 BY SIMILARITY.
 FT DISULFID 1953 1968 BY SIMILARITY.
 FT DISULFID 1971 1986 BY SIMILARITY.
 FT DISULFID 1973 1993 BY SIMILARITY.
 FT DISULFID 1996 2005 BY SIMILARITY.
 FT DISULFID 2008 2022 BY SIMILARITY.
 FT DISULFID 2072 2083 BY SIMILARITY.
 FT DISULFID 2074 2090 BY SIMILARITY.
 FT DISULFID 2092 2101 BY SIMILARITY.
 FT DISULFID 2104 2116 BY SIMILARITY.
 FT DISULFID 2119 2126 BY SIMILARITY.
 FT DISULFID 2121 2133 BY SIMILARITY.
 FT DISULFID 2135 2144 BY SIMILARITY.
 FT DISULFID 2147 2166 BY SIMILARITY.
 FT DISULFID 2169 2169 INTERCHAIN (PROBABLE).
 FT DISULFID 2172 2172 INTERCHAIN (PROBABLE).
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 905 905 N-LINKED (GLCNAC. . .) (POTENTIAL).
 Query Match 6.0%; Score 96; DB 1; Length 3718;
 Best Local Similarity 18.0%; Pred. No. 4.6;
 Matches 76; Conservative 40; Mismatches 96; Indels 210; Gaps 23;
 QY 14 CIDNCGARELKALHLLPOYDDQSFPOSELPKPMTLVGLLPPAKNLNLTQVDNGLP 73
 Db 200 CLERFGPRTLER---ITQDDVI-----CTTEYSRIVPL-ENGEIWSLVNGR-P 244
 QY 74 SAVNGAAPP-----SGPALQGP-----PKITLSGY 98
 Db 245 GAINFSYSPLLRDTKATNIRLRLNTLTLGLHLMKALRDPVTRRYYSIKDISIGR 304
 QY 99 CDCFSSGDFC-----NSC-----SC 113
 Db 305 CVCHGADVCDKPLDPFLRLQACQHTCGGSDRCPCGPNQOPKWPATTDSANECQSC 364
 QY 114 NNLRH-----ELREFKAIKACLDNRNPAFQPKMGKGLGAAL---RHSGKCNKRS 162
 Db 365 NCHGHAYDCYYDPEVDR-----RNASONQDNVYQG--GGVCLDCQHHTTGINGER- 412
 QY 163 GCLKNY-----CEYEAKI---MCSSEI---CK-----CIAC-KNY 190
 Db 413 -CLPGFRAPDPLDSPHVCPCDC-ESDFTDGTCDLTGRCYCRPNFTGELCAACAGY 470
 QY 191 EESPERKMLMSTPHMEPGDFESSHYLSPAKFSPPKLRKNRQAFSCISWEWEATCACL 250
 Db 471 TDFPHCYPLPSFPHN-----DTREQVLP-----GQIVNCDCAA 505
 QY 251 LAQEEAEQE-----HC-----SPLAEQMLEFEGR 277
 Db 506 GTQGNACKRDKPRLGRCKVCKPNFRGAHCELCAPGPHGSPCHPCQCSPPGVANSLCDPESGQ 565
 QY 278 CL 279
 Db 566 CM 567

RESULT 6

KPCL_ASPNG ID KPCL_ASPNG STANDARD; PRT; 1096 AA.
 AC Q00078;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein kinase C-like (EC 2.7.1.-).
 GN PKCA.
 OS Aspergillus niger.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 RN NCBI_TaxID=5061;
 RX SEQUENCE FROM N.A.
 RX MEDLINE=96158841; Pubmed=8569684;
 RA Morawetz R., Lendenfeld T., Mischak H., Muehlbauer M., Gruber F.,
 RA Goodnight J., de Graaff L.H., Visser J., Mushinski J.F.,
 RA Kubicek C.P.;
 RT "Cloning and characterisation of genes (pkc1 and pkcA) encoding
 RT protein kinase C homologues from Trichoderma reesei and Aspergillus
 RT niger";
 RL Mol. Gen. Genet. 250:17-28(1996).
 CC -|- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
 CC BINDING DOMAINS.
 CC -|- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC PKC SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U10549; AAA37433.1; -
 DR HSP; Q63450; 1A06.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR InterPro; IPR000719; Euk_kinase.
 DR InterPro; IPR000961; Pkinase_C.
 DR InterPro; IPR000861; REM_repeat.
 DR InterPro; IPR002230; Ser_thr_kinase.
 DR Pfam; PF00130; DAG_PE-bind; 2.
 DR Pfam; PF02185; HRI; 2.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00433; pkinase_C; 1.
 DR SMART; SM00109; C1; 2.
 DR SMART; SM00239; C2; 1.
 DR SMART; SM00074; HRI; 2.
 DR SMART; SM00133; S_TK_X; 1.
 DR SMART; SM00220; S_TK_X; 1.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 2.
 DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding; zinc;
 KW Phorbol-ester binding; Repeat.
 FT DOMAIN 460 507 PHORBOL-ESTER AND DAG BINDING 1.
 FT DOMAIN 528 577 PHORBOL-ESTER AND DAG BINDING 2.
 FT DOMAIN 771 1030 PROTEIN KINASE.
 FT NP_BIND 777 785 ATP (BY SIMILARITY).
 FT BINDING 800 800 ATP (BY SIMILARITY).
 FT ACT_SITE 896 896 BY SIMILARITY.
 SQ SEQUENCE 1096 AA; 122234 MW; 859B2DB3B3D5B08 CRC64;

Query Match 5.9%; Score 95; DB 1; Length 1096;
 Best Local Similarity 22.8%; Pred. No. 1.5;
 Matches 63; Conservative 21; Mismatches 91; Indels 104; Gaps 15;

QY 66 QVDNGALPSAVN-----GAAPPSPGALQGPPIKTLSCYDCFCSSGDFCSCNNRLHEL 120
 Db 359 KMEHGAPGRPTAGGAGSAHPSGP--GGAAPSGYDGSAGAPAGNQVMIDSW-FAL 415
 QY 121 ERFKAI-----KACLDNRNPEAFQPKMGKGLGAALR-----HSKG-----C 157
 Db 416 EPVGRIYLSMSFAKQLKDRRPF---DIGLNROGAVRQKKEVEHKGQHKFVTQQFYNIM 471
 QY 158 NKRSGCLKNYC---ECYEAKIMCSSIC-----KCIACKNVEESPERKMLMSTPHYMEP 208
 Db 472 RCALCGDFLKYAAGMCQACDKCYTHKKCYPKVVTYTKISKANYETDP----- 517

QY 209 GDFESSHYLSPAKFGPKLRKROAFSCISWEVVEATCACLQAGEAEBOHCSPSLAE 268
 DB 518 -DEEKINRPHRFEG-----FNIS-----ANWC-----HCG----- 545
 QY 269 OMILEEGR-----CLSQILHIEFKSKGLKTE 295
 DB 546 --YLLPFGKNAKRCETGLTCHSHCTHLVDPFCGMSME 582

RESULT 7
 LMA4_HUMAN
 ID LMA4_HUMAN STANDARD; PRT: 1816 AA.
 AC Q16363; Q15335; Q14735; Q9UE18; Q9UJN9;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Laminin alpha-4 chain precursor.
 GN LAMA4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal lung;
 RX MEDLINE=95300971; PubMed=7781776;
 RA Ilvanainen A., Sainio K., Sariola H., Tryggvason K.;
 RT "Primary structure and expression of a novel human laminin alpha 4
 chain.";
 RL FEBS Lett. 365:183-188(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=97454279; PubMed=9310354;
 RA Richards A.J., Luccarini C., Pope F.M.;
 RT "The structural organisation of LAMA4, the gene encoding laminin
 alpha4.";
 RL Eur. J. Biochem. 248:15-23(1997).
 RN [3]
 RP SEQUENCE OF 236-1816 FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=95048381; PubMed=7959779;
 RA Richards A.J., Al-Inara L., Carter N.P., Lloyd J.C., Leversha M.A.,
 RA Pope F.M.;
 RT "Localization of the gene (LAMA4) to chromosome 6q21 and isolation of
 a partial cDNA encoding a variant laminin A chain.";
 RL Genomics 22:237-239(1994).
 RN [4]
 RP SEQUENCE OF 66-1816 FROM N.A.
 RA Tubby B.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 MEMBRANES (MAJOR COMPONENT).
 CC -1- TISSUE SPECIFICITY: IN ADULT, STRONG EXPRESSION IN HEART, LUNG,
 OVARY SMALL AND LARGE INTESTINES, PLACENTA, LIVER; WEAK OR NO
 EXPRESSION IN SKELETAL MUSCLE, KIDNEY, PANCREAS, TESTIS, PROSTATE,
 BRAIN. HIGH EXPRESSION IN FETAL LUNG AND KIDNEY. EXPRESSION IN
 FETAL AND NEWBORN TISSUES IS OBSERVED IN CERTAIN MESENCHYMAL CELLS
 IN TISSUES SUCH AS SMOOTH MUSCLE AND DERMIS.
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAIN G IS GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 3-5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

CC CC -----CAUTION: GENE LAMA4 WAS FORMERLY CALLED LAMA3.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S78569; AAB34635.1; -
 DR EMBL: X91171; CAA62596.1; -
 DR EMBL: Y14240; CAA74636.1; -
 DR EMBL: X76939; CAA54258.1; -
 DR EMBL: Z99289; CAB16553.1; -
 DR HSSP: P02468; IBL0.
 DR MIN: 600133; -
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR001791; Laminin_G.
 DR Pfam: PF00053; laminin_EGF; 3.
 DR Pfam: PF00054; laminin_G; 3.
 DR SMART: SM00180; EGF_Lam; 3.
 DR SMART: SM00282; LamG; 5.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 3.
 DR PROSITE: PS50025; LAM_G_DOMAIN; 5.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 1816 LAMININ ALPHA-4 CHAIN.
 FT DOMAIN 82 131 LAMININ EGF-LIKE 1.
 FT DOMAIN 132 186 LAMININ EGF-LIKE 2.
 FT DOMAIN 187 240 LAMININ EGF-LIKE 3.
 FT DOMAIN 241 255 LAMININ EGF-LIKE 4 (INCOMPLETE).
 FT DOMAIN 256 825 DOMAIN II AND I.
 FT DOMAIN 826 1028 LAMININ G-LIKE 1.
 FT DOMAIN 1040 1220 LAMININ G-LIKE 2.
 FT DOMAIN 1227 1395 LAMININ G-LIKE 3.
 FT DOMAIN 1462 1633 LAMININ G-LIKE 4.
 FT DOMAIN 1640 1813 LAMININ G-LIKE 5.
 FT DOMAIN 313 396 COILED COIL (POTENTIAL).
 FT DOMAIN 466 521 COILED COIL (POTENTIAL).
 FT DOMAIN 574 607 COILED COIL (POTENTIAL).
 FT DOMAIN 655 717 COILED COIL (POTENTIAL).
 FT DOMAIN 770 799 COILED COIL (POTENTIAL).
 FT SITE 717 719 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 82 91 BY SIMILARITY.
 FT DISULFID 84 98 BY SIMILARITY.
 FT DISULFID 101 110 BY SIMILARITY.
 FT DISULFID 113 129 BY SIMILARITY.
 FT DISULFID 132 146 BY SIMILARITY.
 FT DISULFID 134 155 BY SIMILARITY.
 FT DISULFID 157 166 BY SIMILARITY.
 FT DISULFID 169 184 BY SIMILARITY.
 FT DISULFID 187 202 BY SIMILARITY.
 FT DISULFID 189 209 BY SIMILARITY.
 FT DISULFID 212 221 BY SIMILARITY.
 FT DISULFID 224 238 BY SIMILARITY.
 FT DISULFID 266 266 INTERCHAIN (PROBABLE).
 FT DISULFID 269 269 INTERCHAIN (PROBABLE).
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 308 308 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 524 524 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 550 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 639 639 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 735 735 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 780 780 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 803 803 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1086 1086 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1281 1281 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1359 1359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1411 1411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 143 143 A -> P (IN REF. 1).
FT CONFLICT 178 178 L -> F (IN REF. 1).
FT CONFLICT 265 265 G -> GMDPTIS (IN REF. 4).
FT CONFLICT 276 276 D -> A (IN REF. 4).
FT CONFLICT 491 491 Y -> H (IN REF. 2 AND 3).
FT CONFLICT 1057 1057 T -> P (IN REF. 1).
FT CONFLICT 1110 1112 SGR -> GGP (IN REF. 4).
SQ SEQUENCE 1816 AA; 201908 MW; 04E9AF379A0F4A4D CRC64;

```

```

Query Match 5.8%; Score 93; DB 1; Length 1816;
Best Local Similarity 23.8%; Pred. No. 3.8;
Matches 74; Conservative 32; Mismatches 143; Indels 62; Gaps 18;

```

```

QY 2 VICQLKGQMLCIDNGC-----ARELKALH---LLPQY--DDQSGFPQSELKPKMTTLV 51
DB 80 VPDCNGNSNE-CLDGSYCVHCRNTTGEHCEKCLDGYGDSIRGAPQFCPCP----- 133
QY 52 GRLLVPVPAKLNLTQV--DNGALPSAVN-GAAPPSS-----GPALQGGPKITLSGYCDCFSS 104
DB 134 --CPLPLHANAESCVRKNGKAVRCICNENYAGPNCERCAPGYGNPLLI----- 180
QY 105 GDFCNSCSN-----NLRHE--LERFAKAKCLDRNPEAQ-PYMGKGRGLCAKL-RHSK 155
DB 181 GSTCKKDCSGNDPNLIFEDCVTGQCRNCL-RNTTGFECRCAPGYGDIARIAKCA 239
QY 156 GCNCKRSGCLKNYCEVAKIMCSSIC-KCI-----ACKNYERSPERKMLMST--P 203
DB 240 VCNCGGPGDSVYTGCELEEFEPPTGCKVCWDLTDRLAALSIEGKSGVLSVSGAA 299
QY 204 HYMEPGDFESSHYPKSPGPKLRKNRQAFSCISWEVVEATCACLAAQGEAEQHCSS 263
DB 300 AHRHVNEINATVLLTKLSE----RENOVALRKIQINNAENTMKSLSVDVELVEKENQ 355
QY 264 PSIAEQMILEE 274
DB 356 ASRKQLVQKE 366

```

RESULT 8

```

ITBL_XENLA STANDARD; PRT; 798 AA.
AC P12606;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Integrin beta-1 precursor.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88186829; PubMed=28333505;
RA Desimone D.W., Hynes R.O.;
RT "Xenopus laevis integrins. Structural conservation and evolutionary
RT divergence of integrin beta subunits."
RL J. Biol. Chem. 263:5333-5340(1988).
CC -!- FUNCTION: BETA INTEGRINS ASSOCIATE WITH ALPHA SUBUNITS TO FORM
CC RECEPTOR COMPLEXES THAT RECOGNIZE THE SEQUENCE R-G-D IN A WIDE
CC ARRAY OF LIGANDS (BY SIMILARITY).
CC -!- SUBUNIT: HETERODIMER OF AN ALPHA AND BETA SUBUNIT.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY. THE
CC INTEGRINS BETA-1 AND BETA-1* OF XENOPUS ONLY DIFFER BY 19 AMINO

```

```

CC ACIDS.
CC -!- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; M20140; AAA49889.1; -.
DR PIR; A28193; A28193.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002369; Integrin_B.
DR InterPro; IPR001169; Integrin_beta_C.
DR InterPro; IPR003659; PSI.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00362; Integrin_B; 1.
DR PRINTS; PR01186; INTEGRINB.
DR ProDom; PD001811; Integrin_B; 1.
DR SMART; SM00187; INB; 1.
DR SMART; SM00423; PSI; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00243; INTEGRIN_BETA; 3.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
KW Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;
KW Repeat; Signal; Phosphorylation.
FT SIGNAL 1 21
FT CHAIN 22 798 INTEGRIN BETA-1.
FT DOMAIN 22 727 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 728 751 POTENTIAL.
FT DOMAIN 752 798 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 139 377 VWFA-LIKE.
FT DOMAIN 466 635 4 CYSTEINE-RICH TANDEM REPEATS.
FT REPEAT 466 535 I.
FT REPEAT 516 559 II.
FT REPEAT 560 598 III.
FT REPEAT 599 635 IV.
FT DISULFID 28 464 BY SIMILARITY.
FT DISULFID 36 46 BY SIMILARITY.
FT DISULFID 39 76 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 206 212 BY SIMILARITY.
FT DISULFID 260 300 BY SIMILARITY.
FT DISULFID 400 414 BY SIMILARITY.
FT DISULFID 434 691 BY SIMILARITY.
FT DISULFID 462 466 BY SIMILARITY.
FT DISULFID 477 489 BY SIMILARITY.
FT DISULFID 486 525 BY SIMILARITY.
FT DISULFID 491 500 BY SIMILARITY.
FT DISULFID 502 516 BY SIMILARITY.
FT DISULFID 531 536 BY SIMILARITY.
FT DISULFID 533 568 BY SIMILARITY.
FT DISULFID 538 553 BY SIMILARITY.
FT DISULFID 555 560 BY SIMILARITY.
FT DISULFID 574 579 BY SIMILARITY.
FT DISULFID 576 607 BY SIMILARITY.
FT DISULFID 581 590 BY SIMILARITY.
FT DISULFID 592 599 BY SIMILARITY.
FT DISULFID 613 618 BY SIMILARITY.
FT DISULFID 615 661 BY SIMILARITY.
FT DISULFID 620 630 BY SIMILARITY.
FT DISULFID 633 636 BY SIMILARITY.
FT DISULFID 640 649 BY SIMILARITY.
FT DISULFID 646 723 BY SIMILARITY.
FT DISULFID 665 699 BY SIMILARITY.
FT MOD_RES 783 783 PHOSPHORYLATION (BY SIMILARITY).
FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 416 416 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 798 AA; 88167 MW; 093522509B298531 CRC64;

Query Match
Best Local Similarity 5.8%; Score 92; DB 1; Length 798;
Matches 47; Conservative 14; Mismatches 71; Indels 82; Gaps 12;

QY 95 LSGYCD-----CFSSGD-FNCSSCNLRHLEFRKAIKACLDNRNPEAFQPKMGKRL 146
DB 512 MDAYCRRENSSEICSNNGDCIGQCCKRONPNEVSG-KYCECDN---FNCDRSNGLI 567
QY 147 GAAKLHRSKG-CNCKRSGLKNY-----CEYEAKIMCNS-----ICKIACKNVE 191
DB 568 CG-----GKGICKRCVCECFNPGSGSACDCSDTSTCMKNGQICNGRIGCDGRCK--- 619
QY 192 ESPERKMLMSTPHYMPEGDFESSHVSHPAKFSGPKLRKRNQAFSCISWEVVEATC-ACL 250
DB 620 -----CTDPKFGP-----TCCLCQ 634

QY 251 LAQGEAEQHCSPSLAQMILEFGRLSILH 284
DB 635 TCVGVCAEHKEVCQCFQK-GEQDVCMQCMMH 667

RESULT 9
OC90_HUMAN
AC 002509; STANDARD; PRT; 493 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Otc90in 90 precursor (Otc90) (Phospholipase A2 homolog).
GN OTC90 OR PLA2L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93181166; PubMed=8382789;
RA Feuchter-Nurthy A.E., Freeman J.D., Mager D.L.;
RT "Splicing of a human endogenous retrovirus to a novel phospholipase
RL Nucleic Acids Res. 21:135-143(1993).
CC -!- FUNCTION: IT IS UNLIKELY THAT THIS PROTEIN HAS PHOSPHOLIPASE A2
CC ACTIVITY.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY. CONTAINS 3
CC PA2 TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z14310; CAA78662.1; ALT_INIT.
CC HSP; P00593; 4BP2.
CC MIN; 601658; -.
CC InterPro; IPR001211; PLP_A2.
CC Pfam; PF00068; phoslip; 3.
CC PRINTS; PR00389; PHPLIPASEA2.
CC ProDom; PD000303; PLP_A2; 2.
CC SMART; SM00085; PA2C; 2.
CC PROSITE; PS00118; PA2_HIS; 2.

```

```

DR PROSITE; PS00119; PA2_ASP; 1.
KW Signal; Glycoprotein; Repeat.
FT SIGNAL 1 17 POTENTIAL..
FT CHAIN 18 493 OTOCONIN 90.
FT DOMAIN 76 190 PHOSPHOLIPASE A2-LIKE 1.
FT DOMAIN 321 377 PHOSPHOLIPASE A2-LIKE 2.
FT DOMAIN 389 441 PHOSPHOLIPASE A2-LIKE 3.
FT DISULFID 85 145 BY SIMILARITY.
FT DISULFID 99 190 BY SIMILARITY.
FT DISULFID 101 117 BY SIMILARITY.
FT DISULFID 116 172 BY SIMILARITY.
FT DISULFID 123 165 BY SIMILARITY.
FT DISULFID 132 158 BY SIMILARITY.
FT DISULFID 152 163 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 493 AA; 53338 MW; 59F9AF7C1364A5B7 CRC64;

Query Match
Best Local Similarity 5.7%; Score 91; DB 1; Length 493;
Matches 59; Conservative 50; Mismatches 89; Indels 124; Gaps 14;

QY 2 VICQLKGGAQ-MLCIDNGCARELKA-----LHLLPOYDDQSPFPQSELKPMTTLVGR 53
DB 150 IICESKDNCEHLCTCDKAAIECLARSSLSNLLDTSFCLAQTPETTIKEDLTLLPR 209
QY 54 LLPV-PAKLNLI-----TQVD-----NGALPSAVNGAA----- 80
DB 210 VVPVEPTDTSLTALSGEVAATEADRLITLSKKAGHDQEGVGAARATSPPGSAEIVATR 269
QY 81 -----FPSG-----PALQPPPKITLSGYCDCTSSGDFCNSCNSNNLRHELERK 124
DB 270 VTAKIVTLVPAIGIKSLGLAVSSVENDPEETTEKACDRFT---FLHLSGDNHQMVPOLGE 326
QY 125 AIAKCLDRNPEAFQPKMGKRLGAAKLRHSGKCNKRSQ-----CLKNYCECYE 173
DB 327 MLFCLTSRCPEFE-----SYCYCGQEGRPRDLDRCCLSHHC-CLE 370
QY 174 -----AKIMCS-----SICKCIACKNYEESPERKM 198
DB 371 QVRRGLGCLLERLPWSPVVDHTPKGGGSLCEKLLCACDQTAACMTSASFNSQLKSPS 430
QY 199 LMSTPHYMPEGDFESSHYLSPA 220
DB 431 RLGCPCG--OPAAACEDSLHPVPA 450

RESULT 10
EZHL_HUMAN
ID EZHL_HUMAN STANDARD; PRT; 747 AA.
AC Q92800; O43287; Q14459;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enhancer of zeste homolog 1 (ENX-2).
GN EZHL OR KIAA0388.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97079676; PubMed=8921387;
RA Abel K.J., Brody L.C., Valdes J.M., Erdos M.R., McKinley D.R.,
RA Castilla L.H., Merajver S.D., Couch F.J., Friedman L.S.,
RA Ostermeyer E.A., Lynch E.D., King M.-C., Welch P.L.,
RA Osborne-Lawrence S., Spillman M., Bowcock A.M., Collins F.S.,
RA Weber B.L.;
RT "Characterization of EZHL, a human homolog of Drosophila Enhancer of
RT zeste near BRCAL."
RL Genomics 37:161-171(1996).

```


Db	562	RCK-TQCNTKQPCYLAVERCDPDCLTCGASEHWCKVVSCKNCISIQRLKKHL	---	616
QY	205	YMEPGDFESSHYLSPAKFG	-----PPKKRNQAFSCISWEVVEYATCACLAAQE	255
Db	617	-----LAPSDVAGWGTFIKESVQKN	-----EFISYCGELISQDE	651
RESULT	11			
FBN2_MOUSE				
ID	FBN2_MOUSE	STANDARD;	PRT;	2907 AA.
AC	Q61555;	O63957;		
DT	15-DEC-1998	(Rel. 37, Created)		
DT	15-DEC-1998	(Rel. 37, Last sequence update)		
DT	15-DEC-1998	(Rel. 37, Last annotation update)		
DE	Fibrillin 2 precursor.			
GN	FBN2 OR FBN-2.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata;			
OC	Mammalia; Eutheria; Rodentia;			
OC	Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
NC	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=95263670; PubMed=7744963;			
RA	Zhang H., Hu W., Ramirez F.;			
RA	"Developmental expression of fibrillin genes suggests heterogeneity			
RT	of extracellular microfibrils."			
RT	J. Cell Biol. 129:1165-1176(1995).			
RL	[2]			
RP	SEQUENCE OF 210-317 FROM N.A.			
RP	MEDLINE=94140368; PubMed=8307578;			
RA	Li X., Pereira L., Zhang H., Sanguineti C., Ramirez F., Bonadio J.,			
RA	Francce U.;			
RT	"Fibrillin genes map to regions of conserved mouse/human synteny on			
RT	mouse chromosomes 2 and 18."			
RL	Genomics 18:667-672(1993).			
CC	-1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS			
CC	THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE			
CC	LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.			
CC	-1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING			
CC	EGF-LIKE DOMAINS.			
CC	-1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/			
CC	or send an email to license@isb-sib.ch).			
CC	EMBL; L39790; AAA74908.1; -			
DR	EMBL; S69359; AAC60685.1; -			
DR	HSP; P35555; IEMN.			
DR	MGD; MGI:95490; Fbn2.			
DR	InterPro; IPR00152; Asx_hydroxyl.			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR001881; EGF_Ca.			
DR	InterPro; IPR001438; EGF_II.			
DR	InterPro; IPR002212; Tb.			
DR	InterPro; IPR000822; Znf-C2H2.			
DR	Pfam; PF00008; EGF; 46.			
DR	Pfam; PF00683; Tg; 9.			
DR	PRINTS; PR00010; EGFBL00D.			
DR	SMART; SM00179; EGF_CA; 43.			
DR	SMART; SM00001; EGF-like; 3.			
DR	PROSITE; PS00010; ASX_HYDROXYL; 43.			
DR	PROSITE; PS00022; EGF_1; 2.			
DR	PROSITE; PS01186; EGF_2; 36.			
DR	PROSITE; PS01187; EGF_CA; 43.			
KW	Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;			
FT	Repeat; Signal; Multigene family.			
FT	SIGNAL 1 28			
FT	POTENTIAL.			

FT CHAIN 29 2907 FIBRILLIN 2.
FT DOMAIN 111 142 EGF-LIKE 1, NON-CALCIUM BINDING.
FT DOMAIN 145 176 EGF-LIKE 2, NON-CALCIUM BINDING.
FT DOMAIN 176 208 EGF-LIKE 3, NON-CALCIUM BINDING.
FT DOMAIN 276 317 EGF-LIKE 4, CALCIUM-BINDING.
FT DOMAIN 318 359 EGF-LIKE 5, CALCIUM-BINDING.
FT REPEAT 360 426 TGFBP 1.
FT DOMAIN 487 527 EGF-LIKE 6, NON-CALCIUM BINDING.
FT DOMAIN 528 567 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 568 609 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 610 650 EGF-LIKE 9, CALCIUM-BINDING.
FT DOMAIN 651 691 EGF-LIKE 10, CALCIUM-BINDING.
FT REPEAT 692 760 TGFBP 2.
FT DOMAIN 761 802 EGF-LIKE 11, CALCIUM-BINDING.
FT DOMAIN 803 844 EGF-LIKE 12, CALCIUM-BINDING.
FT DOMAIN 845 883 EGF-LIKE 13, CALCIUM-BINDING.
FT DOMAIN 884 929 EGF-LIKE 14, CALCIUM-BINDING.
FT REPEAT 930 1065 TGFBP 3.
FT DOMAIN 1066 1107 EGF-LIKE 15, CALCIUM-BINDING.
FT DOMAIN 1108 1150 EGF-LIKE 16, CALCIUM-BINDING.
FT DOMAIN 1151 1192 EGF-LIKE 17, CALCIUM-BINDING.
FT DOMAIN 1193 1234 EGF-LIKE 18, CALCIUM-BINDING.
FT DOMAIN 1235 1275 EGF-LIKE 19, CALCIUM-BINDING.
FT DOMAIN 1276 1317 EGF-LIKE 20, CALCIUM-BINDING.
FT DOMAIN 1318 1359 EGF-LIKE 21, CALCIUM-BINDING.
FT DOMAIN 1360 1400 EGF-LIKE 22, CALCIUM-BINDING.
FT DOMAIN 1401 1441 EGF-LIKE 23, CALCIUM-BINDING.
FT DOMAIN 1442 1483 EGF-LIKE 24, CALCIUM-BINDING.
FT DOMAIN 1484 1524 EGF-LIKE 25, CALCIUM-BINDING.
FT DOMAIN 1525 1565 EGF-LIKE 26, CALCIUM-BINDING.
FT REPEAT 1566 1642 TGFBP 4.
FT DOMAIN 1643 1684 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1685 1726 EGF-LIKE 28, CALCIUM-BINDING.
FT REPEAT 1727 1800 TGFBP 5.
FT DOMAIN 1801 1842 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1843 1884 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1885 1926 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1927 1965 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1966 2008 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 2009 2048 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2049 2090 EGF-LIKE 35, CALCIUM-BINDING.
FT REPEAT 2091 2163 TGFBP 6.
FT DOMAIN 2164 2205 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2206 2245 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2246 2286 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2287 2330 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2331 2372 EGF-LIKE 40, CALCIUM-BINDING.
FT REPEAT 2373 2441 TGFBP 7.
FT DOMAIN 2442 2483 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2484 2524 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2525 2563 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2564 2606 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2607 2646 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2647 2687 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2688 2727 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 115 124 BY SIMILARITY.
FT DISULFID 132 130 BY SIMILARITY.
FT DISULFID 139 141 BY SIMILARITY.
FT DISULFID 149 159 BY SIMILARITY.
FT DISULFID 153 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 180 190 BY SIMILARITY.
FT DISULFID 184 196 BY SIMILARITY.
FT DISULFID 198 207 BY SIMILARITY.
FT DISULFID 207 222 BY SIMILARITY.
FT DISULFID 280 301 BY SIMILARITY.
FT DISULFID 287 301 BY SIMILARITY.
FT DISULFID 303 316 BY SIMILARITY.
FT DISULFID 322 334 BY SIMILARITY.
FT DISULFID 329 343 BY SIMILARITY.
FT DISULFID 345 358 BY SIMILARITY.
FT DISULFID 491 503 BY SIMILARITY.
FT DISULFID 498 512 BY SIMILARITY.
FT DISULFID 514 526 BY SIMILARITY.

FT DISULFID 532 542 BY SIMILARITY.
FT DISULFID 537 551 BY SIMILARITY.
FT DISULFID 553 566 BY SIMILARITY.
FT DISULFID 572 584 BY SIMILARITY.
FT DISULFID 579 593 BY SIMILARITY.
FT DISULFID 595 608 BY SIMILARITY.
FT DISULFID 614 625 BY SIMILARITY.
FT DISULFID 620 634 BY SIMILARITY.
FT DISULFID 636 649 BY SIMILARITY.
FT DISULFID 655 666 BY SIMILARITY.
FT DISULFID 661 675 BY SIMILARITY.
FT DISULFID 677 690 BY SIMILARITY.
FT DISULFID 765 777 BY SIMILARITY.
FT DISULFID 778 786 BY SIMILARITY.
FT DISULFID 788 801 BY SIMILARITY.
FT DISULFID 807 819 BY SIMILARITY.
FT DISULFID 814 828 BY SIMILARITY.
FT DISULFID 830 843 BY SIMILARITY.
FT DISULFID 849 859 BY SIMILARITY.
FT DISULFID 854 868 BY SIMILARITY.
FT DISULFID 870 883 BY SIMILARITY.
FT DISULFID 952 964 BY SIMILARITY.
FT DISULFID 959 973 BY SIMILARITY.
FT DISULFID 975 988 BY SIMILARITY.
FT DISULFID 1070 1082 BY SIMILARITY.
FT DISULFID 1077 1091 BY SIMILARITY.
FT DISULFID 1093 1106 BY SIMILARITY.
FT DISULFID 1112 1124 BY SIMILARITY.
FT DISULFID 1119 1133 BY SIMILARITY.
FT DISULFID 1135 1149 BY SIMILARITY.
FT DISULFID 1162 1176 BY SIMILARITY.
FT DISULFID 1178 1191 BY SIMILARITY.
FT DISULFID 1197 1209 BY SIMILARITY.
FT DISULFID 1204 1218 BY SIMILARITY.
FT DISULFID 1220 1233 BY SIMILARITY.
FT DISULFID 1239 1250 BY SIMILARITY.
FT DISULFID 1246 1259 BY SIMILARITY.
FT DISULFID 1261 1274 BY SIMILARITY.
FT DISULFID 1280 1292 BY SIMILARITY.
FT DISULFID 1287 1301 BY SIMILARITY.
FT DISULFID 1303 1316 BY SIMILARITY.
FT DISULFID 1322 1334 BY SIMILARITY.
FT DISULFID 1329 1343 BY SIMILARITY.
FT DISULFID 1345 1358 BY SIMILARITY.
FT DISULFID 1364 1377 BY SIMILARITY.
FT DISULFID 1371 1386 BY SIMILARITY.
FT DISULFID 1388 1399 BY SIMILARITY.
FT DISULFID 1405 1418 BY SIMILARITY.
FT DISULFID 1412 1427 BY SIMILARITY.
FT DISULFID 1429 1440 BY SIMILARITY.
FT DISULFID 1446 1458 BY SIMILARITY.
FT DISULFID 1453 1467 BY SIMILARITY.
FT DISULFID 1469 1482 BY SIMILARITY.
FT DISULFID 1488 1499 BY SIMILARITY.
FT DISULFID 1494 1508 BY SIMILARITY.
FT DISULFID 1510 1523 BY SIMILARITY.
FT DISULFID 1529 1540 BY SIMILARITY.
FT DISULFID 1535 1549 BY SIMILARITY.
FT DISULFID 1551 1564 BY SIMILARITY.
FT DISULFID 1647 1659 BY SIMILARITY.
FT DISULFID 1654 1668 BY SIMILARITY.
FT DISULFID 1670 1683 BY SIMILARITY.
FT DISULFID 1689 1701 BY SIMILARITY.

Query Match 5.7%; Score 91; DB 1; Length 2907;

Best Local Similarity 18.3%; Pred. No. 9.6;

Matches 73; Conservative 36; Mismatches 121; Indels 168; Gaps 21;

QY 4 CQLKGGAG--MCLDNCGARELKHLLPOYDQSSPQSELKPKMTTLVGR----- 53

DB 405 CPVRGSEYRRLCLDG-----LP-----MGPIGSSVSRPGGT--GSNGNGYGP 447

Qy 54 ----LLPVPKALNLTQVDNGALPSAVNGAAPSPGALQGP-----PK 92
Db 448 GTGFLPIPG-----DNGFSP-CVGAGVGAGG--GGPIITGLTILNQTDICKHHAN 496
Qy 93 ITLSGYC-----DCF-----SSGDFPNS-----CSCNNL 116
Db 497 LCLNGRCITPVSRCRCNMGYKQDANGDCIDVDECTSNPCSGDCVNTPGSYCKCH-- 554
Qy 117 RHELEKRAIKACLDNRPEAFQPKMGKRLGAALKHSGKCKRSGCLK----- 166
Db 555 -AGFORTPTKQACIDID-ECIQ-----NGVLCKNGRCVNSDGSFQCIEN 596
Qy 167 -----NYCEYEAKIMC-----KCIACKNYEESPCKMLMSTPHYM 206
Db 597 AGFELTTDGNKNCVDHDECTTTNMCLNGMCINEDSGFKCVCKPGFIFLAPNGVCTDWDCEQ 656
Qy 207 EPGDESSHLS---PAKFSPPKLRKRNQAFSCIS-----WEVVEATCA----- 248
Db 657 TPGICMNGHCINNEGSFRCDPCPPGLAVGVDRVCVDTMHRSTCYGEIKKGVCPVPPGAV 716
Qy 249 ----CLLAGGEAEQEHCSPLAEQMILEEGRCLSQI 282
Db 717 TKSECCANPDYGFGEPCQCPAKNS-AEFHGLCSSGI 753

RESULT 12

LMBL_MOUSE STANDARD; PRT; 1786 AA.
AC P02469;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laminin beta-1 chain precursor (Laminin B1 chain).
GN LAMB1-1 OR LAMB-1
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87147212; PubMed=3493487;
RA Sasaki M., Kato S., Kohno K., Martin G.R., Yamada Y.;
RT "Sequence of the cDNA encoding the laminin B1 chain reveals a
multidomain protein containing cysteine-rich repeats.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:935-939(1987).
RN [2]
RP SEQUENCE OF 1292-1786 FROM N.A.
RX MEDLINE=85051302; PubMed=6209134;
RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
coiled-coil alpha-helix.";
RL EMBO J. 3:2355-2362(1984).
RN [3]
RP SEQUENCE OF 165-172; 539-547 AND 712-719.
RX STRAIN=BAUB/C; TISSUE=Endothelial cells;
RC MEDLINE=97363207; PubMed=9219532;
RA Frieser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R.,
RA Sorokin L.M.;
RT "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of
endothelium.";
RL Eur. J. Biochem. 246:727-735(1997).
CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (BHS LAMININ), LAMININ-
2 (MEROSIN), AND LAMININ-6 (K-LAMININ).
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR

COMPONENT).
-1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
-1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
-1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
EMBL; M15525; AAA39407.1; ALT_INIT.
EMBL; X05212; CAA28839.1; -
PIR; A26413; MMSB1.
HSSP; P02468; IKLO.
MGD; MGI:96743; Lamb1-1.
InterPro; IPR000561; EGF-like.
InterPro; IPR001886; LamNT.
InterPro; IPR002049; Laminin_EGF.
Pfam; PF00053; laminin_EGF; 13.
Pfam; PF00055; laminin_Nterm; 1.
PRINTS; PR00011; EGF_LAMININ.
PRODOM; PD002082; LamNT; 1.
SMART; SM00180; EGF_Lam; 11.
SMART; SM00136; LamNT; 1.
PROSITE; PS00022; EGF_1; 9.
PROSITE; PS01186; EGF_2; 2.
PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 1786 LAMININ BETA-1 CHAIN
FT DOMAIN 271 334 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 335 397 LAMININ EGF-LIKE 1.
FT DOMAIN 398 457 LAMININ EGF-LIKE 2.
FT DOMAIN 458 509 LAMININ EGF-LIKE 3.
FT DOMAIN 510 540 LAMININ EGF-LIKE 4.
FT DOMAIN 541 772 LAMININ EGF-LIKE 5 (INCOMPLETE).
FT DOMAIN 773 820 LAMININ DOMAIN IV.
FT DOMAIN 821 866 LAMININ EGF-LIKE 6.
FT DOMAIN 867 916 LAMININ EGF-LIKE 7.
FT DOMAIN 917 975 LAMININ EGF-LIKE 8.
FT DOMAIN 976 1027 LAMININ EGF-LIKE 9.
FT DOMAIN 1028 1083 LAMININ EGF-LIKE 10.
FT DOMAIN 1084 1131 LAMININ EGF-LIKE 11.
FT DOMAIN 1132 1178 LAMININ EGF-LIKE 12.
FT DOMAIN 1179 1397 LAMININ EGF-LIKE 13.
FT DOMAIN 1398 1430 DOMAIN II.
FT DOMAIN 1431 1786 DOMAIN ALPHA.
FT DOMAIN 1787 1815 COILED COIL (POTENTIAL).
FT DOMAIN 1816 1888 COILED COIL (POTENTIAL).
FT DISULFID 271 280 BY SIMILARITY.
FT DISULFID 273 298 BY SIMILARITY.
FT DISULFID 300 309 BY SIMILARITY.
FT DISULFID 312 332 BY SIMILARITY.
FT DISULFID 335 344 BY SIMILARITY.
FT DISULFID 337 362 BY SIMILARITY.
FT DISULFID 365 374 BY SIMILARITY.
FT DISULFID 377 395 BY SIMILARITY.
FT DISULFID 398 411 BY SIMILARITY.
FT DISULFID 400 426 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 440 455 BY SIMILARITY.
FT DISULFID 458 472 BY SIMILARITY.
FT DISULFID 460 479 BY SIMILARITY.
FT DISULFID 481 490 BY SIMILARITY.
FT DISULFID 493 507 BY SIMILARITY.
FT DISULFID 773 785 BY SIMILARITY.
FT DISULFID 775 792 BY SIMILARITY.

FT	DISULFID	794	803	BY SIMILARITY.
FT	DISULFID	806	818	BY SIMILARITY.
FT	DISULFID	821	833	BY SIMILARITY.
FT	DISULFID	823	840	BY SIMILARITY.
FT	DISULFID	842	851	BY SIMILARITY.
FT	DISULFID	854	864	BY SIMILARITY.
FT	DISULFID	867	876	BY SIMILARITY.
FT	DISULFID	869	883	BY SIMILARITY.
FT	DISULFID	886	895	BY SIMILARITY.
FT	DISULFID	898	914	BY SIMILARITY.
FT	DISULFID	917	933	BY SIMILARITY.
FT	DISULFID	919	944	BY SIMILARITY.
FT	DISULFID	946	955	BY SIMILARITY.
FT	DISULFID	958	973	BY SIMILARITY.
FT	DISULFID	976	990	BY SIMILARITY.
FT	DISULFID	978	997	BY SIMILARITY.
FT	DISULFID	1000	1009	BY SIMILARITY.
FT	DISULFID	1012	1025	BY SIMILARITY.
FT	DISULFID	1084	1096	BY SIMILARITY.
FT	DISULFID	1086	1103	BY SIMILARITY.
FT	DISULFID	1105	1114	BY SIMILARITY.
FT	DISULFID	1117	1129	BY SIMILARITY.
FT	DISULFID	1132	1144	BY SIMILARITY.
FT	DISULFID	1134	1151	BY SIMILARITY.
FT	DISULFID	1153	1162	BY SIMILARITY.
FT	DISULFID	1165	1176	BY SIMILARITY.
FT	DISULFID	1179	1179	INTERCHAIN (PROBABLE).
FT	DISULFID	1182	1182	INTERCHAIN (PROBABLE).
FT	DISULFID	1785	1785	INTERCHAIN (PROBABLE).
FT	CARBOHYD	120	120	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	356	356	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	519	519	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	677	677	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1041	1041	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1195	1195	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1279	1279	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1336	1336	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1343	1343	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1487	1487	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1533	1533	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1542	1542	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1643	1643	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	1531	1534	SGNA -> MEMP (IN REF. 2).
FT	CONFLICT	1749	1749	D -> N (IN REF. 2).
SEQ	SEQUENCE	1786 AA;	196904 MW;	846671B7BF41A474 CRC64;

Query Match 5.6%; Score 89; DB 1; Length 1786;
 Best Local Similarity 19.5%; Pred. NO. 8.3;
 Matches 50; Conservative 32; Mismatches 102; Indels 72; Gaps 12;

QY	82	PSGP-----ALQPPKLTLS-----GY-----CDGESSGDF-----CHSCSCN	114
Db	920	PDGDSGRQFARSCYQDPVTTLQACVDPGYIGSRCDGACGFFGNPSDFGSCQPCQCH	979
QY	115	NURHELERFKAKACLDNRPEAFQPKMGR-----GRGAALRHSGK	156
Db	980	---HNID-----TTDPEACDRTGRLCLKLYHTGDHCLQCYGYIGDALRQDCRK	1027
QY	157	CNKRSGGLKNY-----CECYEAKTMCISICKIACKNVEESPERKMLMSTPHYMPGDF	211
Db	1028	CVCNLYGTVKEHCNSDCHDKATQCCLPNVIG-QNCDRCAPNTWQLASGTGCGPCNC	1086
QY	212	ESSHYLSA--KFGPPKLRKRNQAFSC-----ISEWVEATCCLLAQGEAEQEHCS	264
Db	1087	NAHSGFGSCNEFTGOCOMPGFGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTG	1146
QY	265	SLAEQWILE--EFGRC	278
Db	1147	STGQCVCVGVGSPRC	1162

ITB0_XENLA	STANDARD;	PRT;	798 AA.
ID	ITB0_XENLA		
AC	P12607;		
DT	01-OCT-1989 (Rel. 12, Created)		
DT	01-OCT-1989 (Rel. 12, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	Integrin beta-1* precursor.		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8335;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88186829; PubMed=2833505;		
RA	Desimone D.W., Hynes R.O.;		
RT	"Xenopus laevis Integrins. Structural conservation and evolutionary		
RT	divergence of integrin beta subunits.";		
RL	J. Biol. Chem. 263:5333-5340(1988).		
CC	-!- FUNCTION: BETA INTEGRINS ASSOCIATE WITH ALPHA SUBUNITS TO FORM		
CC	RECEPTOR COMPLEXES THAT RECOGNIZE THE SEQUENCE R-G-D IN A WIDE		
CC	ARRAY OF LIGANDS (BY SIMILARITY).		
CC	-!- SUBUNIT: HETERODIMER OF AN ALPHA AND BETA SUBUNIT.		
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-!- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY. THE		
CC	INTEGRINS BETA-1 AND BETA-1* OF XENOPUS ONLY DIFFER BY 19 AMINO		
CC	ACIDS.		
CC	-!- SIMILARITY: CONTAINS 1 WFA-LIKE DOMAIN.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	PIR; B28193; B28193		
DR	InterPro; IPR000561; EGF-like.		
DR	InterPro; IPR002369; Integrin_B.		
DR	InterPro; IPR001169; Integrin_beta_C.		
DR	InterPro; IPR003659; PSI.		
DR	InterPro; IPR002035; WFA.		
DR	Pfam; PF00362; Integrin_B; 1.		
DR	PRINTS; PR01186; INTEGRINB.		
DR	ProDom; PD001811; Integrin_B; 1.		
DR	SMART; SM00187; INB; 1.		
DR	SMART; SM00423; PSI; 1.		
DR	SMART; SM00327; WFA; 1.		
DR	PROSITE; PS00243; INTEGRIN_BETA; 3.		
DR	PROSITE; PS00222; EGF_1; UNKNOWN_2.		
KW	Integrin; Cell adhesion; Receptor; Transmembrane; Glycoprotein;		
KW	Repeat; Signal; Phosphorylation.		
FT	SIGNAL	1	21
FT	CHAIN	22	798
FT	DOMAIN	22	737
FT	EXTRACELLULAR (POTENTIAL).		
FT	POTENTIAL.		
FT	CYTOPLASMIC (POTENTIAL).		
FT	WFA-LIKE.		
FT	4 CYSTEINE-RICH TANDEM REPEATS.		
FT	I.		
FT	II.		
FT	III.		
FT	IV.		
FT	REPEAT	599	635
FT	REPEAT	28	464
FT	DISULFID	36	46
FT	DISULFID	39	76
FT	DISULFID	49	65
FT	DISULFID	206	212
FT	DISULFID	260	300
FT	DISULFID	400	414
FT	DISULFID	434	691

FT DISULFID 462 466 BY SIMILARITY.
FT DISULFID 477 489 BY SIMILARITY.
FT DISULFID 486 525 BY SIMILARITY.
FT DISULFID 491 500 BY SIMILARITY.
FT DISULFID 502 516 BY SIMILARITY.
FT DISULFID 531 536 BY SIMILARITY.
FT DISULFID 533 568 BY SIMILARITY.
FT DISULFID 538 553 BY SIMILARITY.
FT DISULFID 555 560 BY SIMILARITY.
FT DISULFID 574 579 BY SIMILARITY.
FT DISULFID 576 607 BY SIMILARITY.
FT DISULFID 581 590 BY SIMILARITY.
FT DISULFID 592 599 BY SIMILARITY.
FT DISULFID 613 618 BY SIMILARITY.
FT DISULFID 615 661 BY SIMILARITY.
FT DISULFID 620 630 BY SIMILARITY.
FT DISULFID 633 636 BY SIMILARITY.
FT DISULFID 640 649 BY SIMILARITY.
FT DISULFID 646 723 BY SIMILARITY.
FT DISULFID 665 699 BY SIMILARITY.
FT MOD_RES 783 783 PHOSPHORYLATION (BY SIMILARITY).
FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 362 362 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 416 416 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 520 520 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 798 AA; 88303 MW; ALC45EA3711CF7C7 CRC64;

Query Match 5.5%; Score 88; DB 1; Length 798;
Best Local Similarity 21.1%; Pred. No. 4.2;
Matches 45; Conservative 14; Mismatches 74; Indels 80; Gaps 11;

QY 95 LSGVCD-----CFSSGD-FCNSCSNNLRHELEKATKACLDNRPEAFQPKMGKRL 146
Db 512 MDAYCRNSENSEISNNGDCICGQCVCKRDNPNVEYSG-KYCEDN---FNCDRSNGLI 567
QY 147 GAAKLRSKGCNKRSGCLKNY-----CEYEAKTMCSS-----ICKCIAKNYEE 192
Db 568 CGGK---GVCKRCVCEFPNYSAGDCSDSTCTMAKNQICNGRIGCDGCRK---- 619
QY 193 SPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPPKLRNQAFCISWEVVEATC-ACLL 251
Db 620 -----CTDPKFGP-----TCCLCQT 635
QY 252 AQGEAEQEHCSPLAEQMLLEFGRLSQTLLH 284
Db 636 CVGVCTHEKCEVCQRAFQK-GEKQDVCMEOCMH 667

RESULT 14
LMB1_HUMAN
ID LMB1_HUMAN STANDARD; PRT: 1786 AA.
AC P07942;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Laminin beta-1 chain precursor (Laminin B1 chain).
GN LMB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90368768; PubMed=1975589;
RA Vuolteenaho R., Chow L.T., Tryggvason K.;

RT "Structure of the human laminin B1 chain gene.";
J. Biol. Chem. 265:15611-15616(1990).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87280097; PubMed=3611077;
RA Pikkarainen T., Eddy R., Fukushima Y., Byers M., Shows T.,
Pihlajaniemi T., Saraste M., Tryggvason K.;
RT "Human laminin B1 chain. A multidomain protein with gene (LAMB1)
locus in the q22 region of chromosome 7.";
J. Biol. Chem. 262:10454-10462(1987).
[3]
RN SEQUENCE OF 1276-1709 FROM N.A.
RP MEDLINE=88021029; PubMed=3661559;
RA Jaye M., Modi W.S., Ricca G.A., Mudd R., Chiu I.M., O'Brien S.J.,
Drohan W.N.;
RT "Isolation of a cDNA clone for the human laminin-B1 chain and its
gene localization.";
Am. J. Hum. Genet. 41:605-615(1987).
CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-
2 (MEROSIN), AND LAMININ-6 (K-LAMININ).
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
COMPONENT).
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; M61951; AAA59486.1; JOINED.
CC EMBL; M38147; AAA59486.1; JOINED.
CC EMBL; M61917; AAA59486.1; JOINED.
CC EMBL; M61918; AAA59486.1; JOINED.
CC EMBL; M61921; AAA59486.1; JOINED.
CC EMBL; M61922; AAA59486.1; JOINED.
CC EMBL; M61923; AAA59486.1; JOINED.
CC EMBL; M61924; AAA59486.1; JOINED.
CC EMBL; M61925; AAA59486.1; JOINED.
CC EMBL; M61926; AAA59486.1; JOINED.
CC EMBL; M61927; AAA59486.1; JOINED.
CC EMBL; M61928; AAA59486.1; JOINED.
CC EMBL; M61929; AAA59486.1; JOINED.
CC EMBL; M61930; AAA59486.1; JOINED.
CC EMBL; M61931; AAA59486.1; JOINED.
CC EMBL; M61932; AAA59486.1; JOINED.
CC EMBL; M61933; AAA59486.1; JOINED.
CC EMBL; M61934; AAA59486.1; JOINED.
CC EMBL; M61935; AAA59486.1; JOINED.
CC EMBL; M61936; AAA59486.1; JOINED.
CC EMBL; M61938; AAA59486.1; JOINED.
CC EMBL; M61939; AAA59486.1; JOINED.
CC EMBL; M61940; AAA59486.1; JOINED.
CC EMBL; M61941; AAA59486.1; JOINED.
CC EMBL; M61942; AAA59486.1; JOINED.
CC EMBL; M61943; AAA59486.1; JOINED.
CC EMBL; M61944; AAA59486.1; JOINED.

DR EMBL; M61945; AAA59486.1; JOINED.
DR EMBL; M61946; AAA59486.1; JOINED.
DR EMBL; M61947; AAA59486.1; JOINED.
DR EMBL; M61948; AAA59486.1; JOINED.
DR EMBL; M61949; AAA59486.1; JOINED.
DR EMBL; M61950; AAA59486.1; JOINED.
DR EMBL; M53370; AAA59485.1; JOINED.
DR EMBL; M53378; AAA59485.1; JOINED.
DR EMBL; M53371; AAA59485.1; JOINED.
DR EMBL; M53372; AAA59485.1; JOINED.
DR EMBL; M53373; AAA59485.1; JOINED.
DR EMBL; M53374; AAA59485.1; JOINED.
DR EMBL; M53375; AAA59485.1; JOINED.
DR EMBL; M53376; AAA59485.1; JOINED.
DR EMBL; M53344; AAA59485.1; JOINED.
DR EMBL; M53345; AAA59485.1; JOINED.
DR EMBL; M53346; AAA59485.1; JOINED.
DR EMBL; M53347; AAA59485.1; JOINED.
DR EMBL; M53348; AAA59485.1; JOINED.
DR EMBL; M53349; AAA59485.1; JOINED.
DR EMBL; M53350; AAA59485.1; JOINED.
DR EMBL; M53351; AAA59485.1; JOINED.
DR EMBL; M53352; AAA59485.1; JOINED.
DR EMBL; M53353; AAA59485.1; JOINED.
DR EMBL; M53355; AAA59485.1; JOINED.
DR EMBL; M53356; AAA59485.1; JOINED.
DR EMBL; M53357; AAA59485.1; JOINED.
DR EMBL; M53358; AAA59485.1; JOINED.
DR EMBL; M53359; AAA59485.1; JOINED.
DR EMBL; M53360; AAA59485.1; JOINED.
DR EMBL; M53361; AAA59485.1; JOINED.
DR EMBL; M53362; AAA59485.1; JOINED.
DR EMBL; M53363; AAA59485.1; JOINED.
DR EMBL; M53364; AAA59485.1; JOINED.
DR EMBL; M53366; AAA59485.1; JOINED.
DR EMBL; M53367; AAA59485.1; JOINED.
DR EMBL; M53368; AAA59485.1; JOINED.
DR EMBL; M53369; AAA59485.1; JOINED.
DR EMBL; M61916; AAA59482.1; -
DR EMBL; M20206; AAA59487.1; -
DR PIR; SL3347; MMHUB1.
DR HSP; P02468; IKLO.
DR MIN; 150240; -
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT
DR InterPro; IPR002049; Laminin_EGF.
DR Pfam; PF00053; Laminin_EGF; 13.
DR PRINTS; PR00011; EGF_LAMININ.
DR ProDom; PD002082; LamNT; 1.
DR SMART; SM00180; EGF_Lam; 11.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; 9.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 1786 LAMININ BETA-1 CHAIN.
FT DOMAIN 22 270 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 271 334 LAMININ EGF-LIKE 1.
FT DOMAIN 335 397 LAMININ EGF-LIKE 2.
FT DOMAIN 398 457 LAMININ EGF-LIKE 3.
FT DOMAIN 458 509 LAMININ EGF-LIKE 4.
FT DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).
FT DOMAIN 541 771 LAMININ DOMAIN IV.
FT DOMAIN 773 820 LAMININ EGF-LIKE 6.
FT DOMAIN 821 866 LAMININ EGF-LIKE 7.
FT DOMAIN 867 916 LAMININ EGF-LIKE 8.
FT DOMAIN 917 975 LAMININ EGF-LIKE 9.
FT DOMAIN 976 1027 LAMININ EGF-LIKE 10.
FT DOMAIN 1028 1083 LAMININ EGF-LIKE 11.

FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.
FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.
FT DOMAIN 1179 1397 DOMAIN II.
FT DOMAIN 1398 1430 DOMAIN ALPHA.
FT DOMAIN 1431 1786 DOMAIN I.
FT DOMAIN 1216 1315 COILED COIL (POTENTIAL).
FT DOMAIN 1353 1388 COILED COIL (POTENTIAL).
FT DOMAIN 1442 1781 COILED COIL (POTENTIAL).
FT DISULFID 271 280 BY SIMILARITY.
FT DISULFID 273 298 BY SIMILARITY.
FT DISULFID 300 309 BY SIMILARITY.
FT DISULFID 312 332 BY SIMILARITY.
FT DISULFID 335 344 BY SIMILARITY.
FT DISULFID 337 362 BY SIMILARITY.
FT DISULFID 365 374 BY SIMILARITY.
FT DISULFID 377 395 BY SIMILARITY.
FT DISULFID 398 411 BY SIMILARITY.
FT DISULFID 400 426 BY SIMILARITY.
FT DISULFID 428 437 BY SIMILARITY.
FT DISULFID 440 455 BY SIMILARITY.
FT DISULFID 458 472 BY SIMILARITY.
FT DISULFID 460 479 BY SIMILARITY.
FT DISULFID 481 490 BY SIMILARITY.
FT DISULFID 493 507 BY SIMILARITY.
FT DISULFID 773 785 BY SIMILARITY.
FT DISULFID 775 792 BY SIMILARITY.
FT DISULFID 794 803 BY SIMILARITY.
FT DISULFID 806 818 BY SIMILARITY.
FT DISULFID 821 833 BY SIMILARITY.
FT DISULFID 823 840 BY SIMILARITY.
FT DISULFID 842 851 BY SIMILARITY.
FT DISULFID 854 864 BY SIMILARITY.
FT DISULFID 867 876 BY SIMILARITY.
FT DISULFID 869 883 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 898 914 BY SIMILARITY.
FT DISULFID 917 933 BY SIMILARITY.
FT DISULFID 919 944 BY SIMILARITY.
FT DISULFID 946 955 BY SIMILARITY.

Query Match 5.5%; Score 88; DB 1; Length 1786;
Best Local Similarity 18.4%; Pred. No. 10;
Matches 45; Conservative 36; Mismatches 115; Indels 48; Gaps 8;

Qy 82 PSgp-----ALQPPKITLSCYDCFCSSGDCNSCNCNNLRHLERFKAIKACLDLR 132
Db 920 PDGPDGSRQFARSCYQDPVTIQLACVCPGYIGSRCDCCASGYFONPSEVGSCQPCQCH 979

Qy 133 N-----PEAFQPKMKG-----GRLGAALKRHSKGCNKRSGCLKNY 168
Db 980 NNIDTDPDPEACDKETGRCLKCLYHTEGEHCQPCRFYGGDALRQDCRKCVCNLTGVQEH 1039

Qy 169 -----CECYEAKIMCSSICKTACKNYEESPERRKMLMSTPHYMEFGDPFESSHYLSPA--K 221
Db 1040 CNGSDCQCDKATGQCLCLPNVIG-QNCORCAPNTWQLASGTGCDPCNNAHSPGSCNE 1098

Qy 222 FSGPPKLRKRAQFSC-----ISWEVVEATCACLLAQGEAEAEHCSPSLAEOMILE--E 274
Db 1099 FTGQCQCPMPGGRTCTSECQELFWGDDPVECRACDPRGIETPOCDQSTGCGVCVEGVE 1158

Qy 275 FGRC 278
Db 1159 GPRC 1162

RESULT 15
Y070_NPVAC STANDARD; PRT; 290 AA.
ID Y070_NPVAC
AC P41470;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 34.4 kDa protein in LEF3-IAP2 intergenic region.

OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RX MEDLINE=94303173; PubMed=8030224;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus.";
RL Virology 202:586-605(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L22858; AAA66700.1; -
DR HSP; P45628; 1LIR.
KW Hypothetical protein.
SQ SEQUENCE 290 AA; 34408 MW; CA78BA9C8B5AB997 CRC64;

Query Match 5.4%; Score 87; DB 1; Length 290;
Best Local Similarity 20.3%; Pred. No. 1.7;
Matches 37; Conservative 13; Mismatches 48; Indels 84; Gaps 8;

QY 90 PPKITLSCYDC-----FSSGDFCNSGSCNNLRHELRFKAIAKACLDNRPE----- 135
 ||| : ||| || | | | | : ||:
DB 18 PPKFLRTKYCDACRYRFLPKFSEDEKFCGQCICN-----ICNNPKNIDCPS 62

QY 136 ----AFQPKMGKGRGAAKURHSGC-----NCKRSGCL---KNY 168
 : || : : | | : | | : | | : | |
DB 63 SYISKIRPKKENKEIYITSNKNFTCKNECNOQSNRRCLISYFTNESCKELNCCWFNKC 122

QY 169 CECYE-----AKIMCSSICK---CIAC---KNYEESPERKMLM 200
 | | | : | | | | | | | | | | : | : |
DB 123 YMCLEYKKNLYNVNLYTIDGHCPSFKA VCFSCIARIKTCQVCNQPLPKMYKEOEERLKM 182

QY 201 ST 202
DB 183 QS 184

Search completed: July 11, 2002, 08:22:55
Job time: 449 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:14:21 ; Search time 92.59 Seconds

(without alignments)
551.177 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRCLSQLHIEFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPREMBL19:*
- 2: sp.archaea:*
- 3: sp.bacteria:*
- 4: sp.fungi:*
- 5: sp.human:*
- 6: sp.invertebrate:*
- 7: sp.mammal:*
- 8: sp.mhc:*
- 9: sp.organelle:*
- 10: sp.phage:*
- 11: sp.plant:*
- 12: sp.rodent:*
- 13: sp.virus:*
- 14: sp.vertibrate:*
- 15: sp.unclassified:*
- 16: sp.rvirus:*
- 17: sp.bacteriap:*
- 17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	504	31.5	277	11 Q9D571	Q9D571 mus musculus
2	435	27.2	950	5 Q9V608	Q9V608 drosophila
3	396	24.8	429	5 Q62295	Q62295 caenorhabdi
4	396	24.8	435	5 Q95QD7	Q95QD7 caenorhabdi
5	363	22.7	571	10 Q9SL70	Q9SL70 arabidopsis
6	357	22.3	603	10 Q9SZD1	Q9SZD1 arabidopsis
7	253.5	15.9	356	10 Q9LW71	Q9LW71 arabidopsis
8	251.5	15.7	601	10 Q9CAV1	Q9CAV1 arabidopsis
9	232.5	14.5	896	10 Q9CAV1	Q9CAV1 arabidopsis
10	232	14.5	896	10 Q9CAV1	Q9CAV1 arabidopsis
11	223.5	14.0	609	10 Q9M679	Q9M679 arabidopsis
12	222	13.9	526	10 Q9A12	Q9A12 arabidopsis
13	222	13.9	695	10 Q9LE32	Q9LE32 arabidopsis
14	222	13.9	695	10 Q9L013	Q9L013 arabidopsis
15	186	11.6	243	5 Q9VMQ3	Q9VMQ3 drosophila
16	180.5	11.3	593	10 Q9LUI5	Q9LUI5 arabidopsis

ALIGNMENTS

RESULT 1

ID	Q9D571	PRELIMINARY;	PRT;	277 AA.
AC	Q9D571;			
DT	01-JUN-2001 (Tremblrel. 17, Created)			
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)			
DT	01-JUN-2001 (Tremblrel. 17, Last annotation update)			
DE	ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,			
DE	CLONE:4930509C02, FULL INSERT SEQUENCE.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;			
OX	NCBI_TaxID=10090;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=TESTIS;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikolaide I., Pesole G., Quackenbush J.,			
RA	Schriml L. M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M. F.,			
RA	Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D. A., Kamiya M., Lee N. H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarto P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K. F.,			
RA	Suzuki H., Toyo-oka K., Wang K. H., Weitz C., Whittaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,			
RA	Hayashizaki Y.;			
RT	*Functional annotation of a full-length mouse cDNA collection.*;			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK015732; BAB29949.1; .			
SQ	SEQUENCE 277 AA; 28263 MW; 1731F517A3CE4D43 CRC64;			

Q94DS2 oryza sativ
Q75097 homo sapien
Q95980 homo sapien
Q9ET61 rattus norv
Q9JIZ6 rattus norv
O17514 caenorhabdi
Q62335 caenorhabdi
Q94446 chironomus
Q9W2C6 drosophila
Q9PUU4 ictalurus p
Q23460 caenorhabdi
Q9Y1P7 cryptospori
Q9U55 drosophila
Q9V841 drosophila
Q9W343 drosophila
Q9VQR1 drosophila
Q9NKE0 drosophila
Q9VJ08 drosophila
O44341 halotis ru
Q9IEZ0 caprine art
Q9U2A6 caenorhabdi
Q9ZIV0 helicobacte
P91776 pacifastacu
Q9NPY3 homo sapien
Q00274 homo sapien
Q9VVK7 drosophila
Q76733 drosophila
Q75851 homo sapien
Q9D211 mus musculu

Query Match 31.5%; Score 504; DB 11; Length 277;
Best Local Similarity 100.0%; Pred. No. 5.3e-46;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVTCQLKGAGMCLDNCGARELKALHLLPQYDDQSFQSELPKMTTLVGRLLPVPK 60
|||||
Db 181 MVTCQLKGAGMCLDNCGARELKALHLLPQYDDQSFQSELPKMTTLVGRLLPVPK 240
|||||

QY 61 LNLITQVNGALPSAVNGAAPPSPALQGGPKITLSS 97
|||||
Db 241 LNLITQVNGALPSAVNGAAPPSPALQGGPKITLSS 277
|||||

RESULT 2
QYV608 PRELIMINARY; PRT; 950 AA.
ID Q9V608
AC Q9V608;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
DE CG6061 PROTEIN.
GN CG6061.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Cocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Ran K.H., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,
Wan K.R., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abrill J.F., Agbavani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RL EMBL: AE003818; AAF58365.1;
DR FlyBase: FBgn0033846; CG6061.

SQL SEQUENCE 950 AA; 100021 MW; 627C3EA6B44A0A30 CRC64;

Query Match 27.2%; Score 435; DB 5; Length 950;
Best Local Similarity 38.4%; Pred. No. 6.9e-38;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

QY 31 QYDDQSFQSELPKMTTLVGRLLPVPKLNLIQVDNGALPSAVNGAAPPSPALQGP 90
|||||
Db 681 QLPTEQSTPIKVEPKLPTLPQGVKANVPK--PLFEVLKPPATAAAGAVDLGGMTSRR 738
|||||

QY 91 PKITLSS-----YDCFSFGDFCNSCSC-----NNLRHELERFKAIKACLDNRPEAFQPK 140
|||||
Db 739 KHCNCSKSQLKLYDCDFANGFCQDCTCKDCFNLDYEVERAIRSCLDRNPSAFKPK 798
|||||

QY 141 MGKGRGLGAALKRHSKGCNCKRGCLKNYCEYAEKIMCSCICKIACKNVEESPERKMLM 200
|||||

Db 799 ITAPNSGDMRL-HNKGCKNCKRGCLKNYCEYAEKIPCSICKVCGRNWDKPD-----852
|||||

QY 201 STPHYMEPGDFESSHYLSPAKFSGPPKLRNRQ-----AFSCISWEVVEATCACLAAQGE 255
|||||

Db 853 -----VDMBSLDGLMGVEGQKDK-AKNQNLNENRANIVFTDDVIEATIMCMISRV 903
|||||

QY 256 EAEQEHCSPLAEOMILEEFGRCLSOIL 283
|||||

Db 904 MHEKQNVAVEDMEREYMEEMGESLTOII 931
|||||

RESULT 3
QYV6295 PRELIMINARY; PRT; 429 AA.
ID AC 062295;
DT 01-AUG-1998 (TremBLrel. 07, Created)
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE JC8-6B PROTEIN.
GN JC8-6B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lighting J.N.
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology";
RL Science 282:2012-2018(1998).
DR EMBL: 282274; CAB05228.1;
SQ SEQUENCE 429 AA; 48365 MW; 10601B22681C06D2 CRC64;

Query Match 24.8%; Score 396; DB 5; Length 429;
Best Local Similarity 38.9%; Pred. No. 4e-34;
Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

QY 98 YDCFSFGDFCNSCSC-----NNLRHELERFKAIKACLDNRPEAFQPKMGKRLGAALKR- 152
|||||
Db 182 YDCDFANGFCRCNCKDCNHNIEYDSQSKATQSLERNPNAFKPKIARGITDIER 241
|||||

QY 153 -HSKGCNCKRGCLKNYCEYAEKIMCSCICKIACKN-----YEES----PERKML 199
|||||

Db 242 LHKGCHCKRSGCLKNYCEYAEKIPCTCRCKCGCKGCTETRYMTYKNSGGAVSTNAL 301
|||||

QY 200 MS-----TPHYMEPGDFESSHY---LSPAKFSGPPKLRNRAQFSCISW-----EYV 243
|||||

Db 302 MSLTNASSATPD-SGPGSVVTDHGHGDDYEDMLSHKPKVEMDPRRF---PWYMTDEVV 357
|||||

QY 244 EATCACLAAQGEA-----EOEHCSPLAEOMILEEFGRCLSOIL 283
|||||


```

Db 398 EVGEDASKTPASARHRGCKNCKSLKYCEYQGGVCSINCRCEGCKNAFGRKDGSL 457
QY 190 YEESPKMLMSTPHYMEPDESSHYLSPAKFSGPPKLRKNRQ 233
Db 458 FEDEENET-----SGTPGKTKTQO 477

RESULT 12
Q94AI2 PRELIMINARY; PRT; 526 AA.
AC Q94AI2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PUTATIVE DNA BINDING PROTEIN
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Gibson H.A.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yu G., Yu S., Bowser L., Carninci P., Chen H.,
RA Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Natusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RA "Full Length cDNA of gene MW123.15/AT3g22780 (GI:9279696).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046019; AAK76693.1;
SQ SEQUENCE 526 AA; 57886 MW; C6CD9684D417A9D9 CRC64;

Query Match 13.9%; Score 222; DB 10; Length 526;
Best Local Similarity 31.8%; Pred. No. 3e-15;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YDCFFSSGDFC--NSCSC-----NNLRHELEFRFAKAKCLDRNPEAFQPKM----- 141
Db 245 YCECFAAGVYCIIEPCSCIDCFNKPHEETVLAIRKQIESRNPLAFAPKVI RNADSIMEAS 304
QY 142 GKGRGLGAALKRHSKGCNCKRGCKLKYCEYAKIMCSSICKCIACKN----- 189
Db 305 DDASKTPASARHRGCKNCKSLKYCEYQGGVCSINCRCEGCKNAFGRKDGSL 364
QY 190 -----YEESPE--RKMLMSTPHYME-----PGDFESSHYLSPAK--FSGPPKLRKNR 232
Db 365 MESKLEENQETYEKRIAKIQHNVEVSKEVEQNPSDDQSTPLPYRHLVHVHQPFLSKNR 423

RESULT 13
Q9LE32 PRELIMINARY; PRT; 695 AA.
AC Q9LE32;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE CXK DOMAIN PROTEIN TS01 (PUTATIVE DNA BINDING PROTEIN).
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. LER;
RA MEDLINE=20233842; PubMed=10769245;
RA Hauser B.A., He J., Park S.O., Gasser C.S.;

Query Match 13.9%; Score 222; DB 10; Length 695;
Best Local Similarity 31.8%; Pred. No. 4.2e-15;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YDCFFSSGDFC--NSCSC-----NNLRHELEFRFAKAKCLDRNPEAFQPKM----- 141
Db 414 YCECFAAGVYCIIEPCSCIDCFNKPHEETVLAIRKQIESRNPLAFAPKVI RNADSIMEAS 473
QY 142 GKGRGLGAALKRHSKGCNCKRGCKLKYCEYAKIMCSSICKCIACKN----- 189
```

```

RT TS01 is a novel protein that modulates cytokinesis and cell expansion
in Arabidopsis.";
RL Development 127:2219-2236(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG; TISSUE=FLOWER;
RX MEDLINE=20233841; PubMed=10769244;
RA Song J.-Y., Leung T., Ehler L.K., Wang C., Liu Z.;
RT "Regulation of meristem organization and cell division by TS01, an
Arabidopsis gene with cysteine-rich repeats.";
RL Development 127:2207-2217(2000).
DR EMBL: AF204059; AAF69124.1;
DR EMBL: AF206324; AAF27433.1;
SQ SEQUENCE 695 AA; 76258 MW; 25C0BA8550F6B5D4 CRC64;

Query Match 13.9%; Score 222; DB 10; Length 695;
Best Local Similarity 31.8%; Pred. No. 4.2e-15;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YDCFFSSGDFC--NSCSC-----NNLRHELEFRFAKAKCLDRNPEAFQPKM----- 141
Db 414 YCECFAAGVYCIIEPCSCIDCFNKPHEETVLAIRKQIESRNPLAFAPKVI RNADSIMEAS 473
QY 142 GKGRGLGAALKRHSKGCNCKRGCKLKYCEYAKIMCSSICKCIACKN----- 189
Db 474 DDASKTPASARHRGCKNCKSLKYCEYQGGVCSINCRCEGCKNAFGRKDGSL 533
QY 190 -----YEESPE--RKMLMSTPHYME-----PGDFESSHYLSPAK--FSGPPKLRKNR 232
Db 534 MESKLEENQETYEKRIAKIQHNVEVSKEVEQNPSDDQSTPLPYRHLVHVHQPFLSKNR 592

RESULT 14
Q9LUI3 PRELIMINARY; PRT; 695 AA.
AC Q9LUI3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE DNA BINDING PROTEIN-LIKE.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20277480; PubMed=10819329;
RA Nakamura Y.;
RL DNA Rés. 7:131-135(2000).
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
features of the regions of 4,504,864 bp covered by sixty pl and TAC
clones.";
DR EMBL: AB022223; BAB01253.1;
SQ SEQUENCE 695 AA; 76276 MW; 34BBA0E450F6BCE1 CRC64;

Query Match 13.9%; Score 222; DB 10; Length 695;
Best Local Similarity 31.8%; Pred. No. 4.2e-15;
Matches 57; Conservative 16; Mismatches 62; Indels 44; Gaps 7;

QY 98 YDCFFSSGDFC--NSCSC-----NNLRHELEFRFAKAKCLDRNPEAFQPKM----- 141
Db 414 YCECFAAGVYCIIEPCSCIDCFNKPHEETVLAIRKQIESRNPLAFAPKVI RNADSIMEAS 473
QY 142 GKGRGLGAALKRHSKGCNCKRGCKLKYCEYAKIMCSSICKCIACKN----- 189
```

```
DB 474 DDASKTPASARHKGRCNCKSNMCKKYCEYGGVGCNMCRCEGCTNVEGRKDGSLVI 533
QY 190 ----YEESEP--RKMLMSTPHYME-----PGDFESSHYLSPAK--FSGPKLRKNR 232
DB 534 MESKLEENQETYEKRIAKTOHNVENSKEVEQNPSSDQPPPLPPYRHLVYVHOPFLSKNR 592

RESULT 15
Q9VMQ3 PRELIMINARY; PRT; 243 AA.
AC Q9VMQ3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CG14016 PROTEIN.
GN CG14016.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Brill J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AF003610; AAF52259.1; -.
DR HSP: F10414; 1BBG.
DR FlyBase: FBgn0031715; CG14016.
SQ SEQUENCE 243 AA; 26179 MW; B3E11FA1D2605934 CRC64;
```

Query Match 11.6%; Score 186; DB 5; Length 243;
Best Local Similarity 25.4%; Pred. No. 8.4e-12;
Matches 59; Conservative 26; Mismatches 61; Indels 86; Gaps 8;

```
QY 132 RNPEAFQPKMGKRLGAALKRHSKGCNCRKSGCLKNVCEYEAKIMCSSICKCIACKNVE 191
DB 7 RSVDRKADGKKGQ--CAGGV---KGCCCKRSQCIKNYCDYQSMATCTKFCRCVCGRNT 62
QY 192 -----ESPER-----KMLM 200
DB 63 VRELVDPNVSAKNSSAVKQKAAAKAAAGIDVQKALQVAASTLALPKALM 122
QY 201 STPHYM-----EPGDFESSHYLSPAKFSGP-----PKLRKN 231
DB 123 TPKYTLVAGKP--PMASSH--INDIPISRTAATAATPARAVKQPAEPPMPVNLIIIPYRHOD 180
QY 232 ROAFSCISWEVVEATCACLAAOGEEAEQEHCSPLAEOMILEEFGCLSOIL 283
DB 181 RDRNLFVQPVNNAALLECMLIQATEAEQGLNELQVCQLVLEEFMRGYNIL 232
```

Search completed: July 11, 2002, 08:22:18
Job time: 477 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:13:36 ; Search time 34.45 Seconds
(without alignments)
209.160 Million cell updates/sec

Title: us-09-743-237-4

Perfect score: 1599

Sequence: 1 MVICQLKGAQMLCIDNCGA.....GRCLSQLHIFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	93.5	5.8	652	2	US-08-751-305-2
2	87.5	5.5	610	1	US-08-365-470-3
3	87.5	5.5	610	3	US-09-209-668-19
4	87.5	5.5	610	4	US-09-009-490A-89
5	86	5.4	418	4	US-08-795-430-13
6	86	5.4	610	6	5217870-2
7	86	5.4	830	1	US-08-110-158-4
8	85.5	5.3	769	2	US-08-789-078-1
9	85.5	5.3	769	2	US-08-752-633-1
10	85.5	5.3	769	2	US-08-476-062A-45
11	85.5	5.3	769	2	US-07-728-215-31
12	85.5	5.3	769	4	US-08-938-085A-31
13	85.5	5.3	769	5	PCT-US95-04886-1
14	85.5	5.3	769	5	PCT-US96-01314-45
15	85	5.3	676	3	US-08-630-172-10
16	85	5.3	676	4	US-09-375-419-10
17	84.5	5.3	696	3	US-08-899-437-23
18	84.5	5.3	696	4	US-09-126-121-23
19	84.5	5.3	720	3	US-08-899-437-6
20	84.5	5.3	720	4	US-09-126-121-6
21	83.5	5.2	350	2	US-08-999-811-4
22	83.5	5.2	350	2	US-08-824-996-2
23	83.5	5.2	350	3	US-09-042-105-4
24	83.5	5.2	350	4	US-08-510-133A-33
25	83.5	5.2	350	4	US-08-585-895-33
26	83.5	5.2	419	2	US-08-999-811-2
27	83.5	5.2	419	3	US-09-042-105-2

28	83.5	5.2	419	3	US-09-042-105-18	Sequence 18, Appl
29	83.5	5.2	419	4	US-08-795-430-8	Sequence 8, Appl
30	83.5	5.2	419	4	US-08-510-133A-35	Sequence 35, Appl
31	83.5	5.2	419	5	PCT-US96-09001-2	Sequence 2, Appl
32	81	5.1	415	4	US-08-795-430-11	Sequence 11, Appl
33	81	5.1	909	4	US-09-013-895A-4	Sequence 4, Appl
34	81	5.1	2476	2	US-08-276-967-2	Sequence 2, Appl
35	80.5	5.0	3224	2	US-08-705-660-34	Sequence 34, Appl
36	80.5	5.0	3224	3	US-08-989-045-34	Sequence 34, Appl
37	80.5	5.0	4544	1	US-08-489-486-52	Sequence 52, Appl
38	80.5	5.0	4544	2	US-08-489-658-52	Sequence 52, Appl
39	80	5.0	273	1	US-08-152-019A-30	Sequence 30, Appl
40	80	5.0	453	4	US-09-086-483A-5	Sequence 5, Appl
41	80	5.0	1480	3	US-09-191-647-7	Sequence 7, Appl
42	80	5.0	1480	4	US-09-540-245A-7	Sequence 7, Appl
43	80	5.0	1480	4	US-09-540-153-7	Sequence 7, Appl
44	80	5.0	1480	4	US-09-182-024A-5	Sequence 5, Appl
45	80	5.0	1480	5	PCT-US91-09055-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-751-305-2
; Sequence 2, Application US/08751305
; Patent No. 5965439
; GENERAL INFORMATION:
; APPLICANT: Fenner et al., Andrea J.
; TITLE OF INVENTION: HOST DEFENSE ENHANCEMENT
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08751.305
; FILING DATE: 18-NOV-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Wetherell, Jr., John R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: 07306/012001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-305-2

Query Match 5.8%; Score 93.5; DB 2; Length 652;
Best Local Similarity 27.0%; Pred.No.0.57;
Matches 30; Conservative 11; Mismatches 51; Indels 19; Gaps 4;

Qy	96	SGYC--DCFSDDFCNSCNLRHLELFRKAIKACLDNRNPEAFQPKMGKRLGAALRH	153
Db	268	NGCCHQDCEGGDGSFLGCC---RFGFRLLDLVTCAARNPCSSSPCRG-----	313
Qy	154	SKGCNCRSGCLKNY-CECYEAKIMCSSICKCIACKNYEESPERKMLMSTP	203

Db 314 --GATCVLGPBGKNTCTCPQGYQLDSSOLDVDCVDECDSPCAQECVNTVP 362

RESULT 2

US-08-365-470-3
; Sequence 3, Application US/08365470
; Patent No. 5632991
; GENERAL INFORMATION:
; APPLICANT: Gimbrone, Jr., Michael A.
; TITLE OF INVENTION: Antibodies Specific For E-selectin And The Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/365,470
; FILING DATE: herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,510
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/850,802
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Markowicz, Karen R.
; REGISTRATION NUMBER: 36,351
; REFERENCE/DOCKET NUMBER: 0627.1350003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-365-470-3

Query Match 5.5%; Score 87.5; DB 1; Length 610;
Best Local Similarity 21.6%; Pred. No. 2;
Matches 74; Conservative 33; Mismatches 119; Indels 117; Gaps 20;

QY 15 IDNCGARE-----LKALHLLPOYDDOS-----SFPQSELPKPMTTL-----VGR-LLPVPA 59
DB 177 IVNCTALESPEHGLVCSHPLGNFSYNSCSISCDRGYLPSSMETMQCMSGWSADIPA 236
QY 60 ----KLNLITQVDNGALPSAVNGAAPP-----SGPALQGPPIKTLISGYDCDFSSG 105
DB 237 CNVVECDAVTNPANGFVECFQNGSPFWNTCTFDCGEGFELMGAQSL-----QCTSSG 290
QY 106 DF-----CNSCSCNNLRHELERFKAIAKACLDNRNPEAFQPMKGKGRIG-----AAKLRSKG 156
DB 291 NWDNEKPTCAVTCRAVR-----QPONGSVRCSHSPAGEFTFKSS 330
QY 157 CN--CKRSGLKNYCEYEAKIMCSS-----ICKCIACKNYEESPERKMLMSTPHY 205
DB 331 CNFTCEEGFMLOG-----PAQVECTQGTQIPVCEAFQCTAL-SNPER-----GY 377
QY 206 ME--PGDFESSHYLSPAKFSGPP-----KLKRNQAFSCISWEVVEATCACLIAQ----- 253
DB 378 MNCLPSAGSFRYSGSCFSCFCEQGVFLKSKRLQCGGTGEWDNEKPTCEAVRCDVAHQPP 437

QY 254 -----GEEAQEHCSPLAEQOMILEEFG-----RCLSQ 281
DB 438 KGLVRCAHSPIGEFTYKSSCAFSCEEG--FELYGSTQLECTSQ 478

RESULT 3

US-09-209-668-19
; Sequence 19, Application US/09209668A
; Patent No. 6114517
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Xu, Xiaoxing S.
; TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR
; TITLE OF INVENTION: alpha-INDUCED EXPRESSION OF CELL ADHESION MOLECULES
; FILE REFERENCE: ISPH-0336
; CURRENT APPLICATION NUMBER: US/09/209,668A
; CURRENT FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 19
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-668-19

Query Match 5.5%; Score 87.5; DB 3; Length 610;
Best Local Similarity 21.6%; Pred. No. 2;
Matches 74; Conservative 33; Mismatches 119; Indels 117; Gaps 20;

QY 15 IDNCGARE-----LKALHLLPOYDDOS-----SFPQSELPKPMTTL-----VGR-LLPVPA 59
DB 177 IVNCTALESPEHGLVCSHPLGNFSYNSCSISCDRGYLPSSMETMQCMSGWSADIPA 236
QY 60 ----KLNLITQVDNGALPSAVNGAAPP-----SGPALQGPPIKTLISGYDCDFSSG 105
DB 237 CNVVECDAVTNPANGFVECFQNGSPFWNTCTFDCGEGFELMGAQSL-----QCTSSG 290
QY 106 DF-----CNSCSCNNLRHELERFKAIAKACLDNRNPEAFQPMKGKGRIG-----AAKLRSKG 156
DB 291 NWDNEKPTCAVTCRAVR-----QPONGSVRCSHSPAGEFTFKSS 330
QY 157 CN--CKRSGLKNYCEYEAKIMCSS-----ICKCIACKNYEESPERKMLMSTPHY 205
DB 331 CNFTCEEGFMLOG-----PAQVECTQGTQIPVCEAFQCTAL-SNPER-----GY 377
QY 206 ME--PGDFESSHYLSPAKFSGPP-----KLKRNQAFSCISWEVVEATCACLIAQ----- 253
DB 378 MNCLPSAGSFRYSGSCFSCFCEQGVFLKSKRLQCGGTGEWDNEKPTCEAVRCDVAHQPP 437

RESULT 4

US-09-009-490A-89
; Sequence 89, Application US/09009490A
; Patent No. 6300491
; GENERAL INFORMATION:
; APPLICANT: Bennett and Mirabelli
; TITLE OF INVENTION: Oligonucleotide Modulation
; TITLE OF INVENTION: of Cell Adhesion
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Office of Jane Massey Licata
; STREET: 66 East Main Street
; CITY: Marlton
; STATE: NJ
; COUNTRY: USA
; ZIP: 08053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE

[illegible]

Db 280 GPNKELDEETQCVCVGGVRPISCGPHKELDRASCQCMCKNKLPSGCGPNKEFDEEKCQ 339
QY 117 -----RHLEERFKAKACLDNRPEAFQPKMGKGRGLGAALKRHSKGCNCKRSGLKN 167
Db 340 CVCKKTCPRHPLNPAKCTCECTESNKFL-----KG-----KRFHQTCVCRPCTVR 390
QY 168 YCEYEAKIMCISICKI 185
Db 391 TKRCDAFLAEEVCRV 408
RESULT 6
5217870-2
; Patent No. 5217870
; APPLICANT: HESSON, CATHERINE A.; LOBB, ROY R.; GOELZ, SUSAN E.
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES AGAINST CDX
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/345,151
; FILING DATE: 28-APR-1989
; SEQ ID NO: 2:
; LENGTH: 610
5217870-2

Query Match 5.4%; Score 86; DB 6; Length 610;
Best Local Similarity 22.2%; Pred. No. 2.9;
Matches 64; Conservative 29; Mismatches 101; Indels 94; Gaps 17;
QY 15 IDNGARE-----LKAHLHPQYDQGS-----SFPSELKPKMTTL-----VGR-LLPVPA 59
Db 177 IVNCTALESPEHSLVGSHPGLNFSYSSCSISCDRGLPSPMETQCMSSGEWSAIPA 236
QY 60 -----KLNLIITQVNGALPSAVNGAEP-----SGPALOGPKKITLSGYDCDFSSG 105
Db 237 CNVVECDAVTNPANGFVECFQNPFGSPFWNTCTFDCBEGFELMGAQSL-----QCTSSG 290
QY 106 DF-----CNSCSCNNLRHELERFKAKACLDNRPEAFQPKMGKGRGLG-----AAKLRHSG 156
Db 291 NWDNEKPTCKAVTCRAVR-----QPNQSVRCSSHSPAGEFTFKSS 330
QY 157 CN-----CKRSGLKNYCEYEAKIMCSC-----ICKCIACKNYEESPERKMLMSTPHY 205
Db 331 CNFTCEGFMLOG-----PAQVECTQGWTOQIPVCEAFQCTAL-SNPER-----GY 377
QY 206 ME-----PGDFESSHYLSPAKFSPPP-----KLKRNQAFSCISWEVVEATC 247
Db 378 MNCLPSASGSFRYGSCEFSCEQGFVLKSKRLQCGPTGEMDNEKPTC 425

RESULT 7
US-08-110-158-4
; Sequence 4, Application US/08110158
; Patent No. 5605821
; GENERAL INFORMATION:
; APPLICANT: McEver, Rodger P.
; APPLICANT: Pan, Junliang
; TITLE OF INVENTION: Expression Control Sequences of the
; TITLE OF INVENTION: P-Selectin Gene
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 1100 Peachtree Street, Suite 2800
; CITY: Atlanta
; STATE: GA
; COUNTRY: USA
; ZIP: 30309-4530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/110,158
; FILING DATE: 19930820
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/320,408
; FILING DATE: 08-MAR-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404)-815-6508
; TELEFAX: (404)-815-6555
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 830 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-110-158-4

Query Match 5.4%; Score 86; DB 1; Length 830;
Best Local Similarity 22.4%; Pred. No. 4.3;
Matches 46; Conservative 24; Mismatches 79; Indels 56; Gaps 12;
QY 84 GPALOGPKKITLSGYDCDFSSGDF-----CNSCSCNNLRHELERFKAKACLDNRPEAF 137
Db 233 QYQVNGPSKL-----ECLASGIWTKNPKPOCLAAQCPPLKIP-ERGNMI--CL-HSAKAF 282
QY 138 QPKMGKGRGLGAALKRHSKGCNCKRSGLKNYCEYEAKIMCSC-----ICKCIACK 188
Db 283 Q-----HQSACS-----FSCBEGFALVGPVQCTASGVWTAAPYCKAVQCQ 325
QY 189 NYEESPERKMLMSTPHYMGDFESSHYLSPAKFSPPKLR-KNRQAFSCIS-----WEVVE 244
Db 326 HLEAPSEGTMDCVHP-----LTAFYAGSSCKFECPQGYVRGLDMLRCDISGHNSAPL 378
QY 245 ATCACLAAQGEAE-----QEHCSPL 266
Db 379 PTCEAISCEPLESPVHGMDSPSL 403

RESULT 8
US-08-789-078-1
; Sequence 1, Application US/08789078
; Patent No. 5843885
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbets, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; TITLE OF INVENTION: IMMUNE TOLERANCE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/789,078
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/229,513

FILING DATE: 19-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: (816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Tonsil
FEATURE:
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 497..540
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617
OTHER INFORMATION: /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723
OTHER INFORMATION: /label= trans
OTHER INFORMATION: /note= "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769
OTHER INFORMATION: /label= cyto
OTHER INFORMATION: /note= "cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
AUTHORS: Power,
TITLE: LFA-1 Amino acid sequence (B2) (from human
TITLE: tonsil)
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 96-96
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769

US-08-789-078-1

Query Match 5.3%; Score 85.5; DB 2: Length 769;
Best Local Similarity 23.4%; Pred. NO. 4.4;
Matches 32; Conservative 13; Mismatches 37; Indels 55; Gaps 10;

Qy 90 PPKITLSCYDCFS-----SGD-----FCNSCSN-----NLRHELEFRKAIKAC 129
| | | | | : | | | | : | | | : |
Db 525 PKLIYQYCECDTINCERYNQVGGPGRLGICFGKCRCHPFGSGACQGER--TTEGC 582

Qy 130 LDRNPEAFQPKMGKRLGAALKRHSKGNCKRSGLKNVCECYEAKIM-----CSSI 181
| | | | | : | | | | : | | | : |
Db 583 L--NPRRVECS-GRGR-----CRC-----NVCECHSGYQLPLCQECGCTSP 621
| | | | | : | | | | : | | | : |
Qy 182 C-----KCIACKNYEESP 194
| | | | | : | | | | : | | | : |
Db 622 CGKYISCAECLKFEKGP 638
| | | | | : | | | | : | | | : |
RESULT 9:
US-08-752-633-1
; Sequence 1, Application US/08752633
; Patent No. 5863889
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Slahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbetts, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; TITLE OF INVENTION: IMMUNE TOLERANCE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESS: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/752.633
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: (816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..22
; OTHER INFORMATION: /label= signal
; OTHER INFORMATION: /note= "signal sequence"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 449..496
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 497..540
; OTHER INFORMATION: /label= repeat
; OTHER INFORMATION: /note= "cysteine rich repeat"
; FEATURE:
; NAME/KEY: Region

Query Match 5.3%; Score 85.5; DB 2; Length 769;
Best Local Similarity 23.4%; Pred. No. 4.4;
Matches 32; Conservative 13; Mismatches 37; Indels 5

[illegible]

RESULT 11
US-07-728-215-31
; Sequence 31, Application US/07728215
; Patent No. 5962643
; GENERAL INFORMATION.

APPLICANT: Sheppard, Dean
APPLICANT: Quaranta, Vito
APPLICANT: Pytela, Robert
TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States of America
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07728,215
FILING DATE: 19910711
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P31 8717

US-08-938-085A-31

Query Match 5.3%; Score 85.5; DB 4; Length 769;
Best Local Similarity 23.4%; Pred. NO. 4.4;

Qy 90 PPKITLSCYDCFS-----SGD-----FCNSCSN-----NLRHLELRFKAIAK 129
 525 PKLIYGYCECDTINCERYNGOVCGGPGRI/CFGKCRCHPFGSACOCER--TTEGC 582
 Db

Query Match 5.3%; Score 85.5; DB 2; Length 769;
Best Local Similarity 23.4%; Pred. No. 4.4;
Matches 32: Conservative 13; Mismatches 37; Indels 5

Qy	90	PPKITLGYGDCDFS-----SGD-----FCNSCSN-----NLRHLELRFKAIKAC	129
		:	
Db	525	PGKLLIYGVECDTINCERYNGOVCGGPGRGICFCGKCRCHPGFGESACOCFR--TTEGC	582

QY 130 LDNRPEAFQPKWGKGRGLGAALKRLRHSKGNCNCRSGCLKNYCCEYEAKIN-----CSSI 181
| | : | : | | | : | : |
Db 583 L--NPRRVES-GRGR-----CRC-----NVCECHSGYQLPLCQECGCPSP 621

```

QY      182 C---KCIACKNYEESP 194
        |   |   |   |   |
Db      622 CGKYISCAECLKFEKGP 638

```

RESULT 12
US-08-938-085A-31
: Sequence 31, Application US/08938085A
: Patent No. 6339148

/ GENERAL INFORMATION.
 / APPLICANT: Sheppard, Dean
 / APPLICANT: Quaranta, Vito
 / APPLICANT: Pytella, Robert
 / TITLE OF INVENTION: A No. 6339148el Integrin Beta Subunit and Uses
 / TITLE OF INVENTION: Thereof

7
7 CORRESPONDENCE ADDRESS:
7 ADDRESS: Townsend and Crew LLP
7 STREET: Two Embarcadero Center, Eighth Floor
7 CITY: San Francisco
7 STATE: California
7 COUNTRY: USA
7 ZIP: 94111-3834
7

```

:      09-11-80
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/938,085A
: FILING DATE: 26-SEP-1997
:
```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,215
FILING DATE: 11-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Parent, Annette S.
REGISTRATION NUMBER: 42,058
REFERENCE/DOCKET NUMBER: 023070-08021005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

```

; INFORMATION FOR SEQ ID NO: 31:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 769 amino acids
;     TYPE: amino acid
;     STRANDEDNESS:
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
;

```

RESULT 13
PCT-US95-04886-1
; Sequence 1., Application PC/TUS9504886
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbitts, Scott
; TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
; TITLE OF INVENTION: IMMUNE TOLERANCE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.

CITY: Kansas City
 STATE: MO
 COUNTRY: USA
 ZIP: 64106
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/04886

FILING DATE: 01/07/2007
CLASSIFICATION: 28833
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816) 474-9050
TELEFAX: (816) 474-9057
TELEX: 434-363

```

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Tonsil
FEATURE:
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= signal
FEATURE:
NAME/KEY: Region
LOCATION: 449..496

```

```

RESULT 14
PCT-US96-01314-45
; Sequence 45, Application PC/TUS9601314
; GENERAL INFORMATION:
; APPLICANT: M. Amin Arnaout
; TITLE OF INVENTION: METHODS FOR IDENTIFYING
; TITLE OF INVENTION: ANTAGONISTS
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette 1.44 Mb

```

RESULT 15
US-08-630-172-10
; Sequence 10, Application US/08630172
; Patent No. 6060054
; GENERAL INFORMATION:
; APPLICANT: Staerz, Uwe
; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
; TITLE OF INVENTION: LYMPHOCYTE VETO
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, 35th Floor
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,172
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2879-36
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 863-9700
 ; TELEFAX: (303) 863-0223
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 676 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-630-172-10

Query Match 5.38; Score 85; DB 3; Length 676;
 Best Local Similarity 22.18; Pred. No. 4.1;
 Matches 30; Conservative 13; Mismatches 39; Indels 54; Gaps 9;

Qy	90	PPKITLSGYCDGFS-----SGD-----FCNSCSN-----NLRHELERFKAIKAC	129
Db	503	PKLIYGYCECDTINCERYNGVCGPGRLCGKRCRHPGFGSACQGER--TTEGC	560
Qy	130	LDRNPEAFQPKMGKGLGAALKRHSKGCNCRSGCLKNVCECYEAKIN-----C	178
Db	561	L--NPRRVECS-GRGR-----CRC-----NWCECHSGYQLPLCQECGCPGSPC	599
Qy	179	SSICKCIACKNYEESP	194
Db	600	GYISCAECLAKFEKGP	615

Search completed: July 11, 2002, 08:16:00
 Job time: 144 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:13:36 ; Search time 210.79 seconds
(without alignments)
492.596 Million cell updates/sec

Title: US-09-743-237-4

Perfect score: 1599

Sequence: 1 MVICQLKGAQMCLIDNCGA.....GRLSQLHIEFKSKGLKIE 295

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues

Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- Pending_Patents_AA_Main:*
- 1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pap.*
 - 2: /cgn2_6/ptodata/2/paa/US06_COMB.pap.*
 - 3: /cgn2_6/ptodata/2/paa/US07_COMB.pap.*
 - 4: /cgn2_6/ptodata/2/paa/US08_COMB.pap.*
 - 5: /cgn2_6/ptodata/2/paa/US081_COMB.pap.*
 - 6: /cgn2_6/ptodata/2/paa/US082_COMB.pap.*
 - 7: /cgn2_6/ptodata/2/paa/US083_COMB.pap.*
 - 8: /cgn2_6/ptodata/2/paa/US084_COMB.pap.*
 - 9: /cgn2_6/ptodata/2/paa/US085_COMB.pap.*
 - 10: /cgn2_6/ptodata/2/paa/US086_COMB.pap.*
 - 11: /cgn2_6/ptodata/2/paa/US087_COMB.pap.*
 - 12: /cgn2_6/ptodata/2/paa/US088_COMB.pap.*
 - 13: /cgn2_6/ptodata/2/paa/US089_COMB.pap.*
 - 14: /cgn2_6/ptodata/2/paa/US090_COMB.pap.*
 - 15: /cgn2_6/ptodata/2/paa/US091_COMB.pap.*
 - 16: /cgn2_6/ptodata/2/paa/US092_COMB.pap.*
 - 17: /cgn2_6/ptodata/2/paa/US093_COMB.pap.*
 - 18: /cgn2_6/ptodata/2/paa/US094_COMB.pap.*
 - 19: /cgn2_6/ptodata/2/paa/US095_COMB.pap.*
 - 20: /cgn2_6/ptodata/2/paa/US096_COMB.pap.*
 - 21: /cgn2_6/ptodata/2/paa/US097_COMB.pap.*
 - 22: /cgn2_6/ptodata/2/paa/US098_COMB.pap.*
 - 23: /cgn2_6/ptodata/2/paa/US099_COMB.pap.*
 - 24: /cgn2_6/ptodata/2/paa/US100_COMB.pap.*
 - 25: /cgn2_6/ptodata/2/paa/US101_COMB.pap.*
 - 26: /cgn2_6/ptodata/2/paa/US102_COMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1599	100.0	295	21	US-09-743-237-4
2	1599	100.0	295	21	US-09-743-237-24
3	1251	78.2	299	21	US-09-743-237-5
4	505.5	31.6	223	1	PCT-US01-08656-10693
5	435	27.2	403	26	US-60-161-932-1700
6	435	27.2	890	26	US-60-167-217-12948
7	435	27.2	950	20	US-09-614-150-12897

Query Match	100.0%	Score	1599;	DB	21;	Length	295;
Best Local Similarity	100.0%	Pred. No.	1.3e-143;				
Matches	295;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						

Qy 1 MVICQLKGAQMCLIDNCGARELKALHLLPQYDDQSSFPQSELKPMTTLVGRLLPVPAP 60
Db 1 MVICQLKGAQMCLIDNCGARELKALHLLPQYDDQSSFPQSELKPMTTLVGRLLPVPAP 60

ALIGNMENTS

US-09-743-237-4
; Sequence 4, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YUJI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743.237
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-743-237-4

Sequence 12935, A
Sequence 13, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 3366, A
Sequence 3366, A
Sequence 48317, A
Sequence 3367, A
Sequence 2154, Ap
Sequence 2192, Ap
Sequence 2194, A
Sequence 27295, A
Sequence 22002, A
Sequence 22183, A
Sequence 22184, A
Sequence 22002, A
Sequence 22183, A
Sequence 22184, A
Sequence 27296, A
Sequence 27382, A
Sequence 1409, Ap
Sequence 3988, Ap
Sequence 1409, Ap
Sequence 22013, A
Sequence 22013, A
Sequence 21980, A
Sequence 21980, A
Sequence 6922, Ap
Sequence 1881, Ap
Sequence 16915, A
Sequence 32886, A
Sequence 25631, A
Sequence 25631, A
Sequence 25630, A
Sequence 25629, A
Sequence 17053, A
Sequence 17063, A
Sequence 31233, A

US-60-191-637-12935
US-09-087-136-13
US-09-087-136-13
US-09-220-091-13
US-09-708-427-33366
US-09-708-427-33365
PCT-US01-08631-48317
US-09-708-427-33367
US-09-573-655A-2154
US-09-573-655A-2192
US-09-708-427-27294
US-09-708-427-27295
US-09-733-089-22183
US-09-733-089-22184
US-09-816-660-22002
US-09-816-660-22183
US-09-816-660-22184
US-09-708-427-27296
US-60-324-109-27382
PCT-US01-01329-1409
PCT-US01-01329-3988
US-09-764-891-3988
US-10-080-090-1409
US-09-733-089-22013
US-09-816-660-22013
US-09-733-089-21980
US-09-816-660-21980
PCT-US00-26524B-6922
US-09-428-944-1881
US-60-324-109-16915
US-60-324-109-32886
US-09-708-427-25631
US-09-708-427-25630
US-09-708-427-25629
US-60-324-109-17053
US-60-324-109-17063
US-60-324-109-31233

8 435 27.2 950 26
9 396 24.8 438 14
10 396 24.8 438 14
11 396 24.8 438 16
12 363 22.7 550 21
13 363 22.7 571 21
14 362 22.6 280 21
15 357 22.3 524 21
16 357 22.3 603 19
17 357 22.3 603 19
18 357 22.3 603 19
19 352.5 22.0 534 21
20 351.5 22.0 615 21
21 351.5 22.0 615 21
22 351.5 22.0 615 21
23 351.5 22.0 615 22
24 351.5 22.0 615 22
25 351.5 22.0 615 22
26 333.5 20.9 518 21
27 330 20.6 241 26
28 322 20.1 147 1
29 322 20.1 147 1
30 322 20.1 147 1
31 322 20.1 147 24
32 276.5 17.3 207 21
33 276.5 17.3 207 22
34 276 17.3 168 21
35 276 17.3 168 22
36 248 15.3 53 1
37 244 15.3 349 18
38 234.5 14.7 386 26
39 234.5 14.7 556 26
40 232 14.5 497 21
41 232 14.5 542 21
42 232 14.5 658 21
43 230.5 14.4 770 26
44 230.5 14.4 770 26
45 219.5 13.7 402 26

QY 61 LNLITQVNDGALPSAVNGAAPPSPALQGPPIKTLGYCDGFCSSGDFCNSCSCNNLRHEL 120
Db 61 LNLITQVNDGALPSAVNGAAPPSPALQGPPIKTLGYCDGFCSSGDFCNSCSCNNLRHEL 120
QY 121 ERFAIKACLDNRNPEAPQPMKGRGLGAALRHSGKGCNCRSGCLKNYCEYAKIMCSS 180
Db 121 ERFAIKACLDNRNPEAPQPMKGRGLGAALRHSGKGCNCRSGCLKNYCEYAKIMCSS 180
QY 181 IKCKIACKNVEESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRNROAFSCISW 240
Db 181 IKCKIACKNVEESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRNROAFSCISW 240
QY 241 EVBEATCACLQAQEEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295
Db 241 EVBEATCACLQAQEEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295

RESULT 2
US-09-743-237-24
Query Match 100.0%; Score 1599; DB 21; Length 295;
Best Local Similarity 100.0%; Pred. No. 1.3e-143;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
GENERAL INFORMATION:
APPLICANT: SUGIHARA, TAKASHI
APPLICANT: WADHWA, RENU
APPLICANT: KAUL, SUNIL C.
TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
FILE REFERENCE: 084335/0127
CURRENT APPLICATION NUMBER: US/09/743, 237
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: PCT/JP99/03859
PRIOR FILING DATE: 1999-07-16
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 24
LENGTH: 295
TYPE: PRT
ORGANISM: Mus sp.
US-09-743-237-24

QY 1 MVICQLKGGAGMLCIDNCGARELKALHLLPOYDDQSSFFQSELPKPMITLVGRLLPVPAK 60
Db 1 MVICQLKGGAGMLCIDNCGARELKALHLLPOYDDQSSFFQSELPKPMITLVGRLLPVPAK 60
QY 61 LNLITQVNDGALPSAVNGAAPPSPALQGPPIKTLGYCDGFCSSGDFCNSCSCNNLRHEL 120
Db 61 LNLITQVNDGALPSAVNGAAPPSPALQGPPIKTLGYCDGFCSSGDFCNSCSCNNLRHEL 120
QY 121 ERFAIKACLDNRNPEAPQPMKGRGLGAALRHSGKGCNCRSGCLKNYCEYAKIMCSS 180
Db 121 ERFAIKACLDNRNPEAPQPMKGRGLGAALRHSGKGCNCRSGCLKNYCEYAKIMCSS 180
QY 181 IKCKIACKNVEESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRNROAFSCISW 240
Db 181 IKCKIACKNVEESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRNROAFSCISW 240
QY 241 EVBEATCACLQAQEEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295
Db 241 EVBEATCACLQAQEEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295

RESULT 3
US-09-743-237-5
Query Match 100.0%; Score 1599; DB 21; Length 295;
Best Local Similarity 100.0%; Pred. No. 1.3e-143;
Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
GENERAL INFORMATION:
APPLICANT: SUGIHARA, TAKASHI
APPLICANT: WADHWA, RENU
APPLICANT: KAUL, SUNIL C.

APPLICANT: MITSUI, YUJI
TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
FILE REFERENCE: 084335/0127
CURRENT APPLICATION NUMBER: US/09/743, 237
CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: PCT/JP99/03859
PRIOR FILING DATE: 1999-07-16
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 299
TYPE: PRT
ORGANISM: Homo sapiens
US-09-743-237-5
Query Match 78.2%; Score 1251; DB 21; Length 299;
Best Local Similarity 76.3%; Pred. No. 2.2e-110;
Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;
QY 1 MVICQLKGGAGMLCIDNCGARELKALHLLPOYDDQSSFFQSELPKPMITLVGRLLPVPAK 60
Db 1 MVICQLKGGAGMLCIDNCGARELKALHLLPOYDDQSSFFQSELPKPMITLVGRLLPVPAK 60
QY 61 LNLITQVNDGALPSAVNGAAPPSPALQGPPIKTLGYCDGFCSSGDFCNSCSCNNLRHEL 120
Db 61 LNLITQVNDGALPSAVNGAAPPSPALQGPPIKTLGYCDGFCSSGDFCNSCSCNNLRHEL 120
QY 121 ERFAIKACLDNRNPEAPQPMKGRGLGAALRHSGKGCNCRSGCLKNYCEYAKIMCSS 180
Db 121 ERFAIKACLDNRNPEAPQPMKGRGLGAALRHSGKGCNCRSGCLKNYCEYAKIMCSS 180
QY 177 MCSICKIACKNVEESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRNROAFS 236
Db 177 MCSICKIACKNVEESPERKMLMSTHYMEPGDFESSHYLSPAKESGPPKLRNROAFS 236
QY 237 CISWEVVEATCACLQAQEEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295
Db 237 CISWEVVEATCACLQAQEEAEQHCSPSLAEQMILEFGRCLSQILHIEFKSKGLKIE 295

RESULT 4
PCT-US01-08656-10693
Query Match 31.6%; Score 505.5; DB 1; Length 223;
Best Local Similarity 52.4%; Pred. No. 2.1e-39;
Matches 97; Conservative 22; Mismatches 35; Indels 31; Gaps 4;
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc
TITLE OF INVENTION: NOVEL MACROPHAGE NUCLEIC ACIDS AND POLYPEPTIDES
FILE REFERENCE: 21272-066
CURRENT APPLICATION NUMBER: PCT/US01/08656
CURRENT FILING DATE: 2001-04-16
PRIOR APPLICATION NUMBER: 09/522,929
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: 09/770,160
PRIOR FILING DATE: 2001-01-26
NUMBER OF SEQ ID NOS: 10994
SOFTWARE: Custom
SEQ ID NO 10693
LENGTH: 223
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-08656-10693

QY 44 PKMTTTLVGRLLPVPAKLNITQVNDGALPSAVNGAAPPSPALQGPPIKTLG----- 97
Db 44 PKMTTTLVGRLLPVPAKLNITQVNDGALPSAVNGAAPPSPALQGPPIKTLG----- 97
QY 98 -----YCDGFCSSGDFCNSCSC-----NLRHLEFRFAIKACLDNRPE 135
Db 98 -----YCDGFCSSGDFCNSCSC-----NLRHLEFRFAIKACLDNRPE 135

Db 58 ASRPRKPCNCTKSLKLYCDCFANGFCNNCTNCTNLEHENEROKAIKACLDNRPE 117
Qy 136 AFOPKMGRLGAALKRHSKCNCRSGCKNKCCEYEAKEIMCISCKICACKNYEESPE 195
Db 118 AFKPKIGKGESEDRRHSKCNCRSGCKNKCCEYEAKEIMCISCKICACKNYEESPE 177
Qy 196 RKMMLM 200
Db 178 RKTLM 182

RESULT 5
US-60-161-932-1700
; Sequence 1700, Application US/60161932
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: The Drosophila Genome, Primary Nucleic
; TITLE OF INVENTION: Acid Sequences, Systems Containin the Nucleic Acid Sequences
; FILE REFERENCE: CL000122
; CURRENT APPLICATION NUMBER: US/60/161,932
; CURRENT FILING DATE: 1999-10-28
; NUMBER OF SEQ ID NOS: 2626
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1700
; LENGTH: 403
; TYPE: PRT
; ORGANISM: Drosophila
US-60-161-932-1700

Query Match 27.2%; Score 435; DB 26; Length 403;
Best Local Similarity 38.4%; Pred. No. 2.6e-32;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

Qy 31 QYDQSSFPQSELPKPMTTLVGRLLPVPAKLNLTQVDNGALPSAVNGAAPPSPALQGP 90
Db 134 QLPTQSTPIKVEPKLPTLPVGVKANVPAK--PLFEVLKPPATAAAAGAVDPLGWTSSR 191

Qy 91 PKITLSG-----YDCFSDFGNCSC-----NNLRHELERFAKACLDNRPEAFOPK 140
Db 192 KHCNCSQCLKLYCDCFANGFCODCTCKDFNNLDYEVERERAIRSCLDRNPSAFKPK 251

Qy 141 MGKRLGAALKRHSKCNCRSGCKNKCCEYEAKEIMCISCKICACKNYEESPERKMLM 200
Db 252 ITAPNSGDMRL-HKNGCNCRSGCKNKCCEYEAKEIPCSSICKVCGRNMDRDP----- 305

Qy 201 STPHMEPGDFESSHYLSPAKFSGPKLRKNRQ-----AFSCISWEVVEATCACLQAGE 255
Db 306 -----VMDSLDGLMGVEGQKDK-AKNQNLNENRANIYFTDDVIEATIMCMISRI 356

Qy 256 EAEQEHCSPLAEQMILEFRCGLSQIL 283
Db 357 MHEKQNAVDMEREVMEEMGESLTQII 384

RESULT 6
US-60-167-217-12948
; Sequence 12948, Application US/60167217
; GENERAL INFORMATION:
; APPLICANT: Li, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CL000152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12948
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Drosophila

US-60-167-217-12948

Query Match 27.2%; Score 435; DB 26; Length 890;
Best Local Similarity 38.4%; Pred. No. 7.8e-32;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

Qy 31 QYDQSSFPQSELPKPMTTLVGRLLPVPAKLNLTQVDNGALPSAVNGAAPPSPALQGP 90
Db 621 QLPTQSTPIKVEPKLPTLPVGVKANVPAK--PLFEVLKPPATAAAAGAVDPLGWTSSR 678

Qy 91 PKITLSG-----YDCFSDFGNCSC-----NNLRHELERFAKACLDNRPEAFOPK 140
Db 679 KHCNCSQCLKLYCDCFANGFCODCTCKDFNNLDYEVERERAIRSCLDRNPSAFKPK 738

Qy 141 MGKRLGAALKRHSKCNCRSGCKNKCCEYEAKEIMCISCKICACKNYEESPERKMLM 200
Db 739 ITAPNSGDMRL-HKNGCNCRSGCKNKCCEYEAKEIPCSSICKVCGRNMDRDP----- 792

Qy 201 STPHMEPGDFESSHYLSPAKFSGPKLRKNRQ-----AFSCISWEVVEATCACLQAGE 255
Db 793 -----VMDSLDGLMGVEGQKDK-AKNQNLNENRANIYFTDDVIEATIMCMISRI 843

Qy 256 EAEQEHCSPLAEQMILEFRCGLSQIL 283
Db 844 MHEKQNAVDMEREVMEEMGESLTQII 871

RESULT 7
US-09-614-150-12897
; Sequence 12897, Application US/09614150
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12897
; LENGTH: 950
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150-12897

Query Match 27.2%; Score 435; DB 20; Length 950;
Best Local Similarity 38.4%; Pred. No. 8.5e-32;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps 7;

Qy 31 QYDQSSFPQSELPKPMTTLVGRLLPVPAKLNLTQVDNGALPSAVNGAAPPSPALQGP 90
Db 681 QLPTQSTPIKVEPKLPTLPVGVKANVPAK--PLFEVLKPPATAAAAGAVDPLGWTSSR 738

Qy	91	PKITLSC-----YDCFSGSDFCNSC-----NNLRHELERFKATACLDNRNPEAFQPK 140
Db	739	KHCNCSSQCLKLYCDCFANGFCQDCQTKDCFNNDLYEVERERATRSCLDNRNPSAFKPK 798
Qy	141	MKGRLCAALRLRHSKGNCKNRKSLKNYCEYEAKTMCSSICKIACKNYEESPERKMLM 200
Db	799	ITAPNSGDMRL--HNKGCNARSGCLKNYCEYEAKTMCSSICKYCGVRNMDRPD-----852
Qy	201	STPHMEPGDFESSHYLSPAKFSGPKLRKNRQ-----APFCSISWEVETACACLLAAGE 255
Db	853	-----VDMDSLGLMGVEGQKDK-AKNKQLNENRANIYFTDDVIEATIMCMISRTV 903
Qy	256	EAEQEHCSPSLAEQMILEEFGRLCSQL 283
Db	904	MHEKQNTAVEDMREVMEEGMSUTQII 931

```

RESULT      8
US-60-191-637-12935
; Sequence 12935, Application US/60191637
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: NUCLEIC ACID DETECTION KITS COMPRISING
; TITLE OF INVENTION: GENE SEQUENCES EXPRESSED FROM THE DROSOPHILA GENOME, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000392
; CURRENT APPLICATION NUMBER: US/60/191,637
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 42660
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12935
; LENGTH: 950
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-60-191-637-12935

```

```

Query Match      27.2%; Score 435; DB 26; Length 950;
Best Local Similarity 38.4%; Pred. No. 8.5e-32;
Matches 103; Conservative 40; Mismatches 93; Indels 32; Gaps

Qy 31 QYDQSQFPQSELKPKMTTTLGRLLLPVPAKLNKLITQVDNGALPSAVNGAAFPSPGALQCP 90
Db 681 QLPTESQSPIKVEKPLTLPFGVKANPAK--PLFVFLKPPATAAAAGAVPLGGMTSRR 738

Qy 91 PKVILSG-----VCDPSSGDFCNSC-----NNLRHELERFKAICLDRNPEAFQPK 140
Db 739 KHCNCSKSQCLKLYCDCFANGFQDCFCDFCNLDYEVERERAIRSCLDRNPSAFKPK 798

Qy 141 MGKGRGLAALKRHSGCKNCRSGGLKNKYCEYEAKIMCSSICKTACKNYBESPERKMLM 200
Db 799 ITAPNSGDMRL-HNKGCKNCRSGGLKNKYCEYEAKIPCSSICKCVGRNMDRDP----- 852

Qy 201 STPHYMEPGDFESSHYLSPAKFSPPKLRNKRQ-----AFSCISHEVVEATCACLALQGE 255
Db 853 -----VMDSDLGLMGVGGQKKDK-AKKQNLENRANITYFTDDVTEATIMCMISRV 903

Qy 256 EAEQHCSPSLAEQMILEEFGRCLSQIL 283
Db 904 MHEKONVAVEDMREYVMEEMGESLTOII 931

```

```

RESULT          9
US-09-087-136-13
; Sequence 13, Application US/09087136
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202002
; CURRENT APPLICATION NUMBER: US/09/087,136
; CURRENT FILING DATE: 1998-05-28
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28

```

```

; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: prt
; ORGANISM: Caenorhabditis elegans
US-09-087-136-13

```

Query Match 24.8%; Score 396; DB 14; Length 438;
Best Local Similarity 38.9%; Pred. No. 1.5e-28;
Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

Qy	98	YCDCFSSGDFCNCSG----	NNLRHLERFKATKACLDNRPEAFQPMGKGRGLGAALR-	152
Db	191	YCDCFANGFCRDCNCKDCHNNTEYDSQRSKATQSLRNPNAFKPKIGARGITDIER	250	
Qy	153	-HSGGNCCKSGGLKNVCEYEAKIMCSSJCKCIACKN-----YES----	PERKML 199	
Db	251	LHQGCHCKRSGGLKNVCEYEAQVPCTRCKCGCQNTETYMTRYKNSGGAVSTNAL	310	
Qy	200	MS-----TPHYMEPGDFESSHY---	LSPAKFSPGPKLRKNRQAFSCISW-----EYV 243	
Db	311	MSLTNASSTATPD-SGPGSVVTDHGGDDYEDMLLSHKPKVEMDPRF---	PWYMYTDEVV 366	
Qy	244	EATCACLLAAGEBA-----	EQBHCSPSLAEQMIILEFGCLSQL 283	
Db	367	EAATPNCVVAQAEALNAYEKVOTDEKILNNEKVLVLRFGCLRQMI	412	

```

RESULT 10
US-09-087-136-13
; Sequence 13, Application US/09087136A
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202002
; CURRENT APPLICATION NUMBER: US/09/087,136A
; CURRENT FILING DATE: 1998-05-28
; EARLIER APPLICATION NUMBER: 60/047,996
; EARLIER FILING DATE: 1997-05-28
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-087-136-13

```

Query Match	24.8%;	Score 396;	DB 14;	Length 438;
Best Local Similarity	38.9%;	Pred. NO. 1.5e-28;		
Matches 88;	Conservative	32;	Mismatches 62;	Indels 44;
				Gaps 10;

[illegible]

RESULT 11
US-09-220-091-13

```

; Sequence 13, Application US/09220091
; GENERAL INFORMATION:
; APPLICANT: H. Robert Horvitz
; APPLICANT: Craig Ceol
; APPLICANT: Xiaowei Lu
; TITLE OF INVENTION: A TUMOR SUPPRESSOR PATHWAY IN C. ELEGANS
; FILE REFERENCE: 01997/202003
; CURRENT APPLICATION NUMBER: US/09/220,091
; CURRENT FILING DATE: 1998-12-23
; EARLIER FILING DATE: 1998-04-27, 1996
; EARLIER FILING DATE: 1997-05-28
; EARLIER FILING DATE: 1997-05-28
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 438
; TYPE: PRP
; ORGANISM: Caenorhabditis elegans
US-09-220-091-13

Query Match      24.8%; Score 396; DB 16; Length 438;
Best Local Similarity 38.9%; Pred. No. 1.5e-28;
Matches 88; Conservative 32; Mismatches 62; Indels 44; Gaps 10;

QY 98 YDCDFSSGDFCNSCSC-----NNLRHELEFKAIAKACLDNRNPEAFQPKMGKRLGAALKR- 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 191 YDCDFANGECRCDCNCKDCHNTEYDSQSKAIQSLRNPNNAFKIIGITARGITDIER 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 153 -HSKGCNCKRSGLKNCYCEYAKIMCSSICKCIACKN-----YEE-----PERKML 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 251 LHQGCCKKSGCLKNCYCEYAKVPCTDRCKCKGCQCTETRYKNSGGAVSNAL 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 200 MS-----TPHMEFGDRESSHY---LSPAKFSGPPKLRKNQAFSCISW-----EVV 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 311 MSLTNASTATPD-SGPGSVVTDHGDYEDMLLSHRPKVEMDPRP---PWYMTDEV 366

QY 244 EATCACLAGEEA-----EOHCSPLAEQMLLEFGRLCSOIL 283
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 367 EAATMCVAAQEEALNKEVKQTEDEKLINKEKLVLEFGRLCQMI 412

RESULT 12
US-09-708-427-33366
; Sequence 33366, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33366
; LENGTH: 550
; TYPE: PRP
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc_feature
; LOCATION: 1..550
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..550
; OTHER INFORMATION: Ceres Seq. ID 1836305
US-09-708-427-33366

Query Match      22.7%; Score 363; DB 21; Length 550;
Best Local Similarity 32.4%; Pred. No. 3e-25;
Matches 81; Conservative 39; Mismatches 76; Indels 54; Gaps 8;

```

```

QY 36 SSFQSELPKMTTLVGRLL-----PVPKLNLTQVD-NGALPSAVNGAAF 81
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 25 SSMAASIPSPIVTVTRPIITSQAPPTVATIPPPQSQGILHVPIRHPRESNPMR 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 82 PSGALOGPPK-----ITLSGYCDCFSSGDFCNSCSC-----NNLRHELEFKAIAK 127
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 85 PAGETRDGTQKKKQCKNCKHSRCLKYCECFASCTYDCGCNVCNCFNVEPARRQAVE 144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 128 ACLDRNPEAFQPKMGKRLGAALK-----LRHSGCNCKRSGLKNCYCEYAKI 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 145 STLRNPNAPRPKIAASPHGGRDNRNREEVDVVMRLARHKGCHCKKSGCLKYCECFQANI 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 177 MCSSICKCIACKNYEESPERKMLMSTPH-----YME-----PGDFESSHYLSPAKFSGP 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 205 LCSENCKLCKNFESEVRSQSLFPHGESHNLAYLQHANAATGAIGSSGFAS-----APP 260
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 226 PKLRKNQAF 235
|||:|||||:
DB 261 PKRRKGGEIF 270
|||||:|||||:

RESULT 13
US-09-708-427-33365
; Sequence 33365, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33365
; LENGTH: 571
; TYPE: PRP
; ORGANISM: Arabidopsis thaliana
; NAME/KEY: misc_feature
; LOCATION: 1..571
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..571
; OTHER INFORMATION: Ceres Seq. ID 1836304
US-09-708-427-33365

Query Match      22.7%; Score 363; DB 21; Length 571;
Best Local Similarity 32.4%; Pred. No. 3.1e-25;
Matches 81; Conservative 39; Mismatches 76; Indels 54; Gaps 8;

QY 36 SSFQSELPKMTTLVGRLL-----PVPKLNLTQVD-NGALPSAVNGAAF 81
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 46 SSMAASIPSPIVTVTRPIITSQAPPTVATIPPPQSQGILHVPIRHPRESNPMR 105
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 82 PSGALOGPPK-----ITLSGYCDCFSSGDFCNSCSC-----NNLRHELEFKAIAK 127
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 106 PAGETRDGTQKKKQCKNCKHSRCLKYCECFASCTYDCGCNVCNCFNVEPARRQAVE 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 128 ACLDRNPEAFQPKMGKRLGAALK-----LRHSGCNCKRSGLKNCYCEYAKI 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 166 STLRNPNAPRPKIAASPHGGRDNRNREEVDVVMRLARHKGCHCKKSGCLKYCECFQANI 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 177 MCSSICKCIACKNYEESPERKMLMSTPH-----YME-----PGDFESSHYLSPAKFSGP 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 226 LCSENCKLCKNFESEVRSQSLFPHGESHNLAYLQHANAATGAIGSSGFAS-----APP 281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 226 PKLRKNQAF 235
|||:|||||:
DB 282 PKRRKGGEIF 291
|||||:|||||:

RESULT 14

```

```

PCT-US01-08631-48317
; Sequence 48317, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 48317
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-08631-48317

Query Match          22.6%; Score 362; DB 1; Length 280;
Best Local Similarity 57.9%; Pred. No. 1.5e-25;
Matches 66; Conservative 15; Mismatches 23; Indels 10; Gaps 2;

QY 98 YDCFFSGDFCNCSGSC-----NNLRHELERFAKACLDNRNPEAFQPKMGKGLGAALKRH 153
      |||||:||||:| | | | | | | | | | | | | | | | | | : | |
Db   114 YDCFANGFERCNCNTCNYNLEHENERQAKIACLDNRNPEAFKPICKKGESDRRH 173
      |||||:||||| | | | | | | | | | | | | | | | | | : | |

QY 154 SKGCNCRSRGCLKNKYCBCEYAKIMCSICKCIACKNYEESPERRKMLMSTPHYME 207
      |||||:||||| | | | | | | | | | | | | | | | | | : | |
Db   174 SKGCNCRSRGCLKNKYCEYENIIPIWCCR-----YRQILEKAIQLSGAEQLE 221

RESULT 15
US-09-708-427-33367
; Sequence 33367, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33367
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..524
; OTHER INFORMATION: Xaa is any amino acid
; NAME/KEY: misc_feature
; LOCATION: 1..524
; OTHER INFORMATION: Ceres Seq. ID 1836306
US-09-708-427-33367

Query Match          22.3%; Score 357; DB 21; Length 524;
Best Local Similarity 32.2%; Pred. No. 1e-24;
Matches 79; Conservative 39; Mismatches 73; Indels 54; Gaps 8;

QY 41 SELPKMTTLVGRLL-----PVPAKLNLITQVD--NGALPSAVNGAAPSPPGA 86
      :||:| |:| :| | | | | | | | | | | | | | | | | | :| |
Db   4 ASISPVTIVTRPIITSQAAPTATVPITPPPQSOGILHVPIRHPRESNMPRPAGET 63
      :||:| |:| :| | | | | | | | | | | | | | | | | | :| |

QY 87 LQGPPK-----ITLSCYCDCFSSGDFCNCSGSC-----NNLRHELERFAKIAKCLDR 132
      || | | | | | | | | | | | | | | | | | | | | | | :| |
Db   64 RDGPQPQKKCOCKNKRSLKLYECFCFASGYTCDGCNCVNCFNVENEPARQAVESTLER 123
      || | | | | | | | | | | | | | | | | | | | | | | :| |

QY 133 NPFAFPQPMKGRLGAALK-----LRHSKGCNCKRSGLCKLNKYCEYEAKIMCSSI 181

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:13:36 ; Search time 30.19 Seconds
(without alignments)
1014.815 Million cell updates/sec

Title: US-09-743-237-4
Perfect score: 1599
Sequence: 1 MVICQLKGGQAQMLCIDNCGA.....GRCISQILHTFEKSKGLKIE 295

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 323636 seqs, 103855142 residues

Total number of hits satisfying chosen parameters: 323636

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*

- 1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
- 2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
- 3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
- 4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
- 5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
- 6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
- 7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	357	22.3	603	5	US-09-573-655B-2154
2	357	22.3	603	5	US-09-573-655B-2192
3	333.5	20.9	282	6	US-10-155-881-9626
4	314	19.6	510	6	US-10-155-881-9875
5	248	15.5	53	6	US-10-108-698-6932
6	232.5	14.5	896	6	US-10-155-881-20031
7	205	12.8	359	6	US-10-155-881-9183
8	165	10.3	143	6	US-10-155-881-9877
9	159	9.9	120	6	US-10-138-145-1127
10	117.5	7.3	200	6	US-10-155-881-9182
11	115	7.2	1774	1	PCT-US02-01339-6
12	94.5	5.9	538	1	PCT-US02-10812-17
13	94.5	5.9	577	1	PCT-US02-10812-15
14	92.5	5.8	298	6	US-10-155-881-33654
15	92.5	5.8	580	6	US-10-155-881-11197
16	92.5	5.8	621	6	US-10-155-881-11230
17	92.5	5.8	753	6	US-10-155-881-25969
18	92.5	5.8	809	6	US-10-155-881-11231
19	92.5	5.8	1015	6	US-10-155-881-28881
20	91.5	5.7	554	6	US-10-155-881-22871
21	91.5	5.7	732	6	US-10-155-881-34231
22	90	5.6	3298	7	US-60-365-384-181
23	90	5.6	3332	7	US-60-365-384-182
24	89	5.6	574	7	US-60-360-039-5730
25	88.5	5.5	260	6	US-10-155-881-10280
26	87.5	5.5	955	6	US-10-154-675-14

27	87	5.4	501	1	PCT-US02-17382-130	Sequence 130, App
28	87	5.4	867	5	US-09-895-913A-358	Sequence 358, App
29	87	5.4	1743	5	US-09-882-227-624	Sequence 624, App
30	86.5	5.4	326	6	US-10-155-881-33671	Sequence 33671, A
31	86.5	5.4	338	6	US-10-155-881-21921	Sequence 21921, A
32	86	5.4	1036	6	US-10-121-049-142	Sequence 142, App
33	86	5.4	1036	6	US-10-121-050-142	Sequence 142, App
34	86	5.4	1036	6	US-10-121-053-142	Sequence 142, App
35	86	5.4	1036	6	US-10-121-043-142	Sequence 142, App
36	86	5.4	1036	6	US-10-121-044-142	Sequence 142, App
37	86	5.4	1036	6	US-10-121-047-142	Sequence 142, App
38	86	5.4	1036	6	US-10-121-054-142	Sequence 142, App
39	86	5.4	1036	6	US-10-121-056-142	Sequence 142, App
40	86	5.4	1036	6	US-10-121-057-142	Sequence 142, App
41	86	5.4	1036	6	US-10-121-058-142	Sequence 142, App
42	86	5.4	1036	6	US-10-121-060-142	Sequence 142, App
43	86	5.4	1036	6	US-10-121-063-142	Sequence 142, App
44	86	5.4	1036	6	US-10-123-108-142	Sequence 142, App
45	86	5.4	1036	6	US-10-123-154-142	Sequence 142, App

ALIGNMENTS

RESULT 1
US-09-573-655B-2154
; Sequence 2154, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptide
; FILE REFERENCE: 2750-0876P
; CURRENT APPLICATION NUMBER: US/09/573,655B
; CURRENT FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 3281
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2154
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-573-655B-2154

Query Match 22.3%; Score 357; DB 5; Length 603;
Best Local Similarity 35.2%; Pred. No. 8.6e-24;
Matches 88; Conservative 38; Mismatches 66; Indels 58; Gaps 12;

QY	36	SSFPQSELPKMTTLVGRLLVPVPAKLNLTQVDNGALPSAVNGA----	APFSGPALQGP- 90
DB	60	SSVPSTIRP-GMTIAGQVTVRPTLPMTTMSNPSSQIYNAPIRHPIESP	KARGPR 118
QY	91	PKI-----	TLGYCDCFSFGDFCNCSGC-----NLRHLELERPKAIK 127
DB	119	PNVEGRDGPQKKKQCNKHSRCLKLYCECFASGTGDCGNCVNCFNNDNEP	PARREAVE 178
QY	128	ACLDNRPEAFQPKMGK-----	LGAAKL--RHSKGCNCKRSGCLKNKYCECYEAKI 176
DB	179	ATERNPFPAFRPKRIASSPHGGRRDKREDIGEVLLGKHNGKCHCKSGCL	KYCECFQANI 238
QY	177	MCSSTICKCTACKNYEBSPEPKMLMSTPH-----	YMB-----PGDFESSHYL-SPAKFS 223
DB	239	LCSENCKCLDCNFKFEGSEERQALFGEHSNHMAYLQQAANAATGAVGSSG	FAPSPA--- 295
QY	224	GPKLKRKNRQ 233	
DB	296	--PKRRKGQE 303	

RESULT 2
US-09-573-655B-2192
; Sequence 2192, Application US/09573655B
; GENERAL INFORMATION:
; APPLICANT: SOLOVYEV, Victor and TROUKHAN, Maxim

```

Qy 144  GRL-----GAAKLRLHSGCNCNKGSLGCLKNVCECYEAKIMCSSICKCIACKN 189
      |                                     :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 132  SPLMHRNNEVPSDPLVGV---KHNKGCHCKSGCLKKYCECFQANILCSENCKMCDKN 187
      |                                     :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Qy 190  YEESPERKMLMSTPH 204
      | | | | | | | | | |
Db 188  FEGSEERNLFQGDH 202
      | | | | | | | | | |

RESULT 4
US-10-155-881-9875
: Sequence 9875, Application US/10155881

```

APPLICANT: Dossoli, Stancon B.
APPLICANT: Kovalic, David K.
APPLICANT: Liu, Jingdong
APPLICANT: Lutfiyya, Linda L.
APPLICANT:

; AFFILIANT: MCARENCH, James
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
 ; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
 ; FILE REFERENCE: 38-21(15300)J

; CURRENT APPLICATION NUMBER: 05/10/1557001
 ; CURRENT FILING DATE: 2002-05-22
 ; NUMBER OF SEQ ID NOS: 37595
 ; SEQ ID NO 9875

; ENROLL: J10
 ; TYPE: PRT
 ; ORGANISM: Zea mays
 US-10-155-881-9875

Query Match	19.6%;	Score 314;	DB 6;	Length 510;
Best Local Similarity	35.6%;	Pred. No. 4.9e-20;		

Matches 72, conservative 20, highmatches 40, indels 50, gaps 60

Qy 82 PGGPALQ-GPPKITLSG-----YCDFSSGDFCNCS 112
||| | : |||||
||| | : |||||

```

Qy 113 C-----NNLRHELFRFAKACILDRNPFAQPMKMGKGL-----CA--AKLRHSGKGN 153
Db 112 CKQCGNTVENEKGRQEAINTTKLRNPNAFQPKITENGPISSVWKMDAGALPSHPKHNKGCH 171
Qy 159 CKRSGCLKNTCCECYEAKIMCSICKCIACKNYBESPE-----RKMLLMSTPHYHM 206
Db 172 CKSGCLKLYKCECFQANILCSKNCKMCDKNYEGSELRLTTQGDNSCDNRNNTQQAAANVA 231
Qy 207 EPGDFESSHY-LGPAKFSGGPK 227
Db 232 LNGAIGSSGYRFSVPVRKRPE 253

RESULT 5
US-10-106-698-6932
; Sequence 6932, Application US/10106698
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides a
; FILE REFERENCE: PA005P1

```

CURRENT FILING DATE: 2002-03-27
 PRIOR APPLICATION NUMBER: PCT/US00/26524
 PRIOR FILING DATE: 2000-09-28
 PRIOR APPLICATION NUMBER: US 60/157,137

; PRIOR FILING DATE: 1999-09-23
 ; PRIOR APPLICATION NUMBER: US 60/163,280
 ; PRIOR FILING DATE: 1999-11-03
 ; NUMBER OF SEQ ID NOS: 8564

```

; SOFTWARE: FACULTY VCL 3.0
; SEQ ID NO 6932
; LENGTH: 53
; TYPE: PRT

```

ORGANISM: homo sapiens

US-10-106-698-6932

Query Match 15.5%; Score 248; DB 6; Length 53;
Best Local Similarity 90.6%; Pred. No. 2.4e-15;
Matches 48; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 243 VEATCACLAAQEEAEQEHCSPLAEQMLIEFGRLCSQILHIEFKSKGLKIE 295
DB 1 VEATCACLAAQEEAEKEHCSKCLAEQMLIEFGRLCSQILHIEFKSKGLKME 53

RESULT 6

US-10-155-881-20031
; Sequence 20031, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155.881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 20031
; LENGTH: 896
; TYPE: PRT
; ORGANISM: Glycine max
US-10-155-881-20031

Query Match 14.5%; Score 232.5; DB 6; Length 896;
Best Local Similarity 38.3%; Pred. No. 1.9e-12;
Matches 44; Conservative 16; Mismatches 38; Indels 17; Gaps 3;

QY 98 YDCDFSSGDFC--NSCSNNLRHELEFRKAI-----KACLDNRPEAFQPKM----- 141
DB 487 YDCDFAAAGTYCDPCACOGCLNRPEYETVETKQOIESRNPDAFAPKIVQPTDISHM 546

QY 142 -CKGRGLGAALKRHSKGCNCKRSGCLKNYCEYEAQIMGSSICKIACKNYESPE 195
DB 547 DDENLTTFSARHRRKSGCNKRWCLKKYCEYQANVGCGSGCRCEGCKNVHCKKE 601

RESULT 7

US-10-155-881-9183
; Sequence 9183, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155.881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9183
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9183

Query Match 12.8%; Score 205; DB 6; Length 359;
Best Local Similarity 24.5%; Pred. No. 1.7e-10;
Matches 64; Conservative 31; Mismatches 90; Indels 76; Gaps 10;

QY 56 PVPAKL---NLITQVDNGA--LPSAVNGAAPPSPGALOGPPKI----- 93
DB 18 PPPATLEPGDLPTAFAGGTELEPSSNMQLAIAPDPK---RPRVEAADNGCKHCKACKS 74
QY 94 -TLSGYCDCFSSGDFC--NSCSNNLRHE-----LERFKATKACLDNRPEAF 137
DB 75 RCLKLYPCFSGGGYCDKCGQPCFNKEAFATVHTTRKVLRSQKRMKMKINRRPEAN 134
QY 138 QPKMGKGRGLGAALKRHSKGCNCKRSGCLKNYCEYEAQIMGSSICKIACKN----- 189
DB 135 TEPMEDAHSSSSTPPKRCNCKKSKLKYCDYODGTGCLFRCDDCCQNPFGKNEGI 194
QY 190 -----YVESPERKMLMSTPHYMEPGDFESSHYLSPAKFSGPKLR--KNR 232
DB 195 MADDKRYLYTCADLDHSEGEHDFVVERSRLQSPISKESFHOI-----PPLRASSR 248

QY 233 QA-----FSCISWEVVEATCAC 249
DB 249 DAHVPPHAISQWQALPRSWHC 269

RESULT 8

US-10-155-881-9877
; Sequence 9877, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155.881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9877
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9877

Query Match 10.3%; Score 165; DB 6; Length 143;
Best Local Similarity 38.5%; Pred. No. 2.1e-07;
Matches 35; Conservative 16; Mismatches 24; Indels 16; Gaps 4;

QY 95 LSGYCDCFSSGDFCNSCSNN---LRHELEFRKATKACLDNRPEAFQPKMG----- 142
DB 52 LKMYCECFQELQYCDGCSNCGNIVGNENARNEAIEAIRQNRPSAFQPKIGNDSNTLNV 111

QY 143 -KGRGLGAALKRHS--KGCNCKRSGCLKNYCE 170
DB 112 RKDKFWSSPSRSKPKHCKACKS-CLKKYCD 141

RESULT 9

US-10-138-145-1127
; Sequence 1127, Application US/10138145
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: Gibson, John Bryan
; APPLICANT: Norriss, Michael Geoffrey
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use.
; FILE REFERENCE: 11000.1058U
; CURRENT APPLICATION NUMBER: US/10/138.145
; CURRENT FILING DATE: 2002-05-01
; NUMBER OF SEQ ID NOS: 1652
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1127
LENGTH: 120
TYPE: PRT
ORGANISM: Lolium perenne
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(120)
OTHER INFORMATION: Xaa = Any Amino Acid
US-10-138-145-1127

Query Match 9.9%; Score 159; DB 6; Length 120;
Best Local Similarity 38.7%; Pred. No. 5.9e-07;
Matches 41; Conservative 13; Mismatches 34; Indels 18; Gaps 7;
QY 98 YDCFSGGDFCSC---NLRHELRFKAKACLDNRPEAFQPKMGKGRGLGA-----148
Db 16 YCEC-RAGVYCEPCSCQCGCLNKPHEEIVLSTRKQIEFRNPLA-APKVIRLSDAAQETQE 73
QY 149 -----AKLRHSGCKNCRSGCLKNYCEYEAQIMCSICKCIACKN 189
Db 74 DPNTPASARHKRCNCKKSSCLK-YCEYQGGVGLTNCRC-ECKN 117

RESULT 10
US-10-155-881-9182
Sequence 9182, Application US/10155881
GENERAL INFORMATION:
APPLICANT: Kovalic, Stanton B.
APPLICANT: Kovalic, David K.
APPLICANT: Liu, Jingdong
APPLICANT: Lutfiyya, Linda L.
APPLICANT: McIninch, James
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
TITLE OF INVENTION: TRANSCRIPTION IN PLANTS
FILE REFERENCE: 38-21(15300)J
CURRENT APPLICATION NUMBER: US/10/155,881
CURRENT FILING DATE: 2002-05-22
NUMBER OF SEQ ID NOS: 37595
SEQ ID NO 9182
LENGTH: 200
TYPE: PRT
ORGANISM: Zea mays
US-10-155-881-9182

Query Match 7.3%; Score 117.5; DB 6; Length 200;
Best Local Similarity 24.7%; Pred. No. 0.0056;
Matches 39; Conservative 15; Mismatches 57; Indels 47; Gaps 4;
QY 39 PQSELKPMPTLVG-----RLLPVP--AKNLITQVDNGALPSAVNGAAPSPGALQGGP 91
Db 26 PQAAAPVMDADVAGATNQWPEPDDAHLSPPPPPPPASAPTAAEAATAAASAAVSP 85
QY 92 KITLSGYCDPSSGDFCNSCSNLRHELRFKAKACLDNRPEAFQPKMGKGRGLGA 151
Db 86 -----APAEANGNSDRKRRKADDDGG-- 107
QY 152 RHSGCKNCRSGCLKNYCEYEAQIMCSICKCIACKN 189
Db 108 --CRACCKSRCLKLYCVYASGSHCTELCGCEPCHN 143

RESULT 11
PCT-US02-01339-6
Sequence 6, Application PC/TUS0201339
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: LEE, Ernestine A.
APPLICANT: WALIA, Narinder K.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: AZIMZAI, Yalda
APPLICANT: TANG, Y. Tom

APPLICANT: YUE, Henry
APPLICANT: THANGAVELU, Kavitha
APPLICANT: XU, Yuming
APPLICANT: ARVIZU, Chandra
APPLICANT: WARREN, Bridget A.
APPLICANT: YAO, Monique G.
APPLICANT: AU-YOUNG, Janice
APPLICANT: HAFALIA, April J.A.
APPLICANT: ELLIOTT, Vicki S.
APPLICANT: KALLICK, Deborah A.
APPLICANT: GANDHI, Ameena R.
APPLICANT: RICHARDSON, Thomas W.
APPLICANT: KHAN, Farrah A.
APPLICANT: LU, Yan
APPLICANT: SWARNAKAR, Anita
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: NGUYEN, Dannel B.
APPLICANT: GRAUL, Richard
APPLICANT: LU, Dying Aina M.
TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS
FILE REFERENCE: PI-0346 PCT
CURRENT APPLICATION NUMBER: PCT/US02/01339
CURRENT FILING DATE: 2002-01-16
PRIOR APPLICATION NUMBER: 60/262,838; 60/265,927; 60/271,196; 60/274,549; 60/334,1
PRIOR FILING DATE: 2001-01-19; 2001-02-02; 2001-02-23; 2001-03-09; 2001-11-28
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PERL Program
SEQ ID NO 6
LENGTH: 1774
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 7722591CD1
PCT-US02-01339-6

Query Match 7.2%; Score 115; DB 1; Length 1774;
Best Local Similarity 23.1%; Pred. No. 0.13;
Matches 73; Conservative 27; Mismatches 124; Indels 92; Gaps 18;
QY 17 NCGARELKALHLLPOYDDQSSFFQSELPKPMPTLVGRLLPVPKAKNLITQVDNGALPSAV 76
Db 748 NCNA-----WLLPDLTRSAS-----VGPPEESVAHAAVAGSRLYISGGFGGVALGRL 796
QY 77 NGAAPSGPA--LQGPPIITLGYCD-CFS---SGD-----FCNSCSCNLL---RHELER 122
Db 797 LALTPDPDCRLISSPEACNOSGACTWCHGACLSGDDQAHRLGCGGSPCPMPSPSPCECRR 856
QY 123 FKAIKACLDNRPEAFQPKMGKGRGLGAALKRHSKGC-NCKRSGCL-----KNYC----- 169
Db 857 LRTCSECLARHPTLQPDGE-----ASTPRCKWCTNCPGACIGRNGSCTSENDCRINQ 911
QY 170 -----ECYEAKI-----MCSICKCIACKNYEESPE-RKMLMSTPHY----- 205
Db 912 REVFWAGNCSEAGGAADCEQTRGKCMWTRQPKRTGTRRILSVQPTDWTCTFSHLL 971
QY 206 -MEPGDFESSHYLSPKFSPPKLRKNRQAFSCISWEVVEATCACLGAQGEAEQHCSP 264
Db 972 NVSPMPVRESS---PPLPCPTPCHELLPN-----CYS-----CLDSKGADGGQHCWV 1014
QY 265 SLAEQMILEEFGRCLS 280
Db 1015 SSSLQ-----QCLS 1023

RESULT 12
PCT-US02-10812-17
Sequence 17, Application PC/TUS0210812
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: GANDHI, Ameena R.
APPLICANT: DELEGEANE, Angelo M.

APPLICANT: SWARNAKAR, Anita
APPLICANT: HAFALIA, April J.A.
APPLICANT: DUGGAN, Brendan M.
APPLICANT: WARREN, Bridget A.
APPLICANT: EMERLING, Brooke M.
APPLICANT: ARVIZU, Chandra S.
APPLICANT: HONCHELL, Cynthia D.
APPLICANT: KALLICK, Deborah A.
APPLICANT: LU, Dyung Aina M.
APPLICANT: LEE, Ernestine A.
APPLICANT: YUE, Henry
APPLICANT: FORSYTHE, Ian J.
APPLICANT: RAMKUMAR, Javalaxmi
APPLICANT: GRIFFIN, Jennifer A.
APPLICANT: LI, Joana X.
APPLICANT: THANGAVELU, Kavitha
APPLICANT: BAUGHN, Mariah R.
APPLICANT: YAO, Monique G.
APPLICANT: SANJANWALA, Madhu M.
APPLICANT: WALIA, Narinder K.
APPLICANT: BURFORD, Neil
APPLICANT: LAL, Preeti, G.
APPLICANT: BECHA, Shanya D.
APPLICANT: LEE, Soo Y.
APPLICANT: ELLIOTT, Vicki S.
APPLICANT: LUO, Wen
APPLICANT: LU, Yan
APPLICANT: WANG, Yu-mei E.
TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
FILE REFERENCE: PI-0397 PCT
CURRENT APPLICATION NUMBER: PCT/US02/10812
CURRENT FILING DATE: 2002-04-08
PRIOR APPLICATION NUMBER: 60/282,282; 60/283,782; 60/284,823; 60/287,264;
60/288,662; 60/290,383; 60/298,348; 60/351,928;
60/359,903
PRIOR FILING DATE: 2001-04-05; 2001-04-13; 2001-04-18; 2001-04-27;
2001-05-04; 2001-05-11; 2001-06-15; 2002-01-25;
2002-02-25
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PERL Program
SEQ ID NO 17
LENGTH: 538
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 7506402CD1
PCT-US02-10812-17

Query Match 5.9%; Score 94.5; DB 1; Length 538;
Best Local Similarity 20.5%; Pred. No. 2.1;
Matches 62; Conservative 37; Mismatches 110; Indels 93; Gaps 17;

QY 14 CIDNCGARELKAHLHPQYDQSSFPQSELPKPMTTLVGRLLPVPKLNLTQVDNGALP 73
DB 76 CYSNCSPPKFTQILML-----FPPNLYPKET-----LEAFVIVTQM----- 112

QY 74 SAVNGAAPPSPGALQGPVKITLSGYDCFSFGDFCN-----SCSCNNLRHEL 120
DB 113 -----LALSLSGISYDDPKK-----CQSESTCINNPVQSGVKGTSFSSCSLSRFSQNF 161

QY 121 ERFAKAKCLDRNPAFPQKMGKGRGLRAKLRHSGCKCK-----RSGLKN 167
DB 162 SNV-GVK-CLQNKPO-MQKSPKPCVGNRLEGNEICDGTAEQCGPASCDFRTCVLKD 218

QY 168 YCECYEAKIMCSSICKC-----TACKNYEESPERKMLMSTPPHYMEPGDRSSHVLSPAKFS 223
DB 219 GAKCY--KGLCKDCQIQLQSGVECRP-KAHPECDTA-----ENCNGSSPE--C 261

QY 224 GPP-----KLKNRQAFSCISNEV--VEATCACLLAQGEAEQHCSPSLAQOMILEEFG 276
DB 262 GPDITLNLGCKNKKFYCYDGDCHDLDAKCESVFGKSGSRNAPPACYEIEIOSQS--DREG 319

QY 277 RC 278
DB 320 NC 321

RESULT 13

PCT-US02-10812-15
; Sequence 15, Application PC/TUS0210812
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: GANDHI, Ameen R.
; APPLICANT: DELEGEANE, Angelo M.
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: WARREN, Bridget A.
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: HONCHELL, Cynthia D.
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: YUE, Henry
; APPLICANT: FORSYTHE, Ian J.
; APPLICANT: RAMKUMAR, Javalaxmi
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: LI, Joana X.
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: YAO, Monique G.
; APPLICANT: SANJANWALA, Madhu M.
; APPLICANT: WALIA, Narinder K.
; APPLICANT: BURFORD, Neil
; APPLICANT: LAL, Preeti, G.
; APPLICANT: BECHA, Shanya D.
; APPLICANT: LEE, Soo Y.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: LUO, Wen
; APPLICANT: LU, Yan
; APPLICANT: WANG, Yu-mei E.
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PI-0397 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/10812
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 60/282,282; 60/283,782; 60/284,823; 60/287,264;
60/288,662; 60/290,383; 60/298,348; 60/351,928;
60/359,903
; PRIOR FILING DATE: 2001-04-05; 2001-04-13; 2001-04-18; 2001-04-27;
2001-05-04; 2001-05-11; 2001-06-15; 2002-01-25;
2002-02-25
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PERL Program
; SEQ ID NO 15
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 55064352CD1
PCT-US02-10812-15

Query Match 5.9%; Score 94.5; DB 1; Length 577;
Best Local Similarity 20.5%; Pred. No. 2.3;
Matches 62; Conservative 37; Mismatches 110; Indels 93; Gaps 17;

QY 14 CIDNCGARELKAHLHPQYDQSSFPQSELPKPMTTLVGRLLPVPKLNLTQVDNGALP 73
DB 76 CYSNCSPPKFTQILML-----FPPNLYPKET-----LEAFVIVTQM----- 112
QY 74 SAVNGAAPPSPGALQGPVKITLSGYDCFSFGDFCN-----SCSCNNLRHEL 120

Db 113 -----LALSIGISYDDPKK-----COCSESTCMMPEVVQSNVKTFSFSSCSLRSTQFI 161
QY 121 ERFKAIAKALDRNPEAFQPKMGKRLGAAKLRHSGKCNCK-----RSGCLKN 167
Db 162 SNV-GVK-CLQNKPO-WQKSPKPCVCGNGLENEICDCGTEAQCGPASCDDPRTCVLKD 218
QY 168 YCEYKAKIMCSICKC-----IACKNYEESPERKMLMSTPHYMEPGDFESSHYLSPAKFS 223
Db 219 GAKCY--KGLCCCKDQILQSGVECRP-KAHPECDIA-----ENCNGSSPE--C 261
QY 224 GPP-----KLKRNQAFSCISWEV--VEATCACLQAQEEAEQHCSPSLAEQMIILEFG 276
Db 262 GPDTITLGLSKNNKFCYDGDCHLDARCESVFGKGRNAPFACYEEIQSOS--DRFG 319
QY 277 RC 278
Db 320 NC 321

RESULT 14
US-10-155-881-33654
; Sequence 33654, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 33654
; LENGTH: 298
; TYPE: PRT
; ORGANISM: zea mays
US-10-155-881-33654

Query Match 5.8%; Score 92.5; DB 6; Length 298;
Best Local Similarity 23.7%; Pred. No. 1.6; Mismatches 29; Indels 49; Gaps 7;
Matches 27; Conservative 9; Mismatches 29; Indels 49; Gaps 7;
QY 94 TLSGYDCFSFGDFCNS-CSCNNLRHELRFKAIAKALDRNPEAFQPKMGKRLGAAKLR 152
Db 53 TCGKQCPCLSNGTCCCKYCGC-----PKICKNRF----- 81
QY 153 HSKGCNKRSGCLKNKYCEYKIMCS-SICK-C-IAC-----KNYE 191
Db 82 --RGCHCAKSQCRSQCPCFAADRECDPDVCRNCWVGCGDGTGLGVPNQRGDNYE 133

RESULT 15
US-10-155-881-11197
; Sequence 11197, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 11197
; LENGTH: 580
; TYPE: PRT
; ORGANISM: zea mays

US-10-155-881-11197
Query Match 5.8%; Score 92.5; DB 6; Length 580;
Best Local Similarity 23.7%; Pred. No. 3.5; Mismatches 27; Conservative 9; Mismatches 29; Indels 49; Gaps 7;
Matches 27; Conservative 9; Mismatches 29; Indels 49; Gaps 7;
QY 94 TLSGYDCFSFGDFCNS-CSCNNLRHELRFKAIAKALDRNPEAFQPKMGKRLGAAKLR 152
Db 335 TCGKQCPCLSNGTCCCKYCGC-----PKICKNRF----- 363
QY 153 HSKGCNKRSGCLKNKYCEYKIMCS-SICK-C-IAC-----KNYE 191
Db 364 --RGCHCAKSQCRSQCPCFAADRECDPDVCRNCWVGCGDGTGLGVPNQRGDNYE 415
Search completed: July 11, 2002, 08:14:15
Job time: 39 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:16:59 ; Search time 53.06 Seconds
(without alignments)
625.915 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647

Sequence: 1 MVICQLKGGTQMLCIDSRT.....GRCLSQLHTFEKSKGLKME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_032802:*

1: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1647	100.0	299	21	AA198464 Human testis speci
2	1251	76.0	295	21	AA198463 Mouse testis speci
3	470	28.5	950	22	ABB62035 Drosophila melanog
4	446	27.1	147	22	AA195330 Human reproductive
5	441	26.8	438	20	AA198392 Caenorhabditis ele
6	378	23.0	280	22	ABG17958 Novel human diagno
7	278	16.9	53	22	AA1976158 Human colon cancer
8	216	13.1	251	22	AA1993348 Human polypeptide,
9	200	12.1	243	22	ABB68888 Drosophila melanog
10	113	6.9	1981	22	ABB61657 Drosophila melanog
11	109.5	6.6	699	20	AA1981840 Human LFA-1 beta c

12 109.5 6.6 769 9 AAP80836 Beta subunit of hu

13 109.5 6.6 769 11 RAR07113 Recombinant beta-s

14 109.5 6.6 769 16 AAB0108 LFA-1 beta subunit

15 109.5 6.6 769 21 AAB03974 LFA-1 CD18 polyep

16 109.5 6.6 793 22 ABG16399 Novel human diagno

17 108.5 6.6 769 13 AAR24256 Beta-subunit CD18

18 108 6.6 676 18 AAR35853 Human CD18 for use

19 108 6.6 3011 13 AAR21519 Compilled HCV seque

20 105.5 6.4 695 22 ABB30979 Peptide #3630 enco

21 105.5 6.4 695 22 ABB36170 Peptide #3676 enco

22 105.5 6.4 695 22 AAM56953 Human brain expres

23 105.5 6.4 695 22 AAM69338 Human bone marrow

24 105.5 6.4 695 22 AAM17170 Peptide #3604 enco

25 105.5 6.4 695 22 AAM29661 Peptide #3698 enco

26 105.5 6.4 695 22 AAM04863 Peptide #3545 enco

27 104 6.3 3011 14 AAR31621 Hepatitis C virus

28 103.5 6.3 1725 21 AAB19800 Mouse laminin 2 ma

29 103.5 6.3 1725 21 AAB48451 Mouse laminin 8 po

30 103.5 6.3 1764 10 AAP91672 Primary amino acid

31 103.5 6.3 1776 19 AAW50894 Mouse laminin B1 c

32 103.5 6.3 1786 21 AAB19799 Mouse laminin 2 be

33 103.5 6.3 1786 21 AAB48450 Drosophila melanog

34 102 6.2 1788 22 ABB62995 Novel human diagno

35 101.5 6.2 762 22 ABB08032 HCV J1 NS3-NS4 dom

36 101 6.1 507 16 AAR66631 Hepatitis C virus

37 101 6.1 609 15 AAR51170 A nonstructural pr

38 101 6.1 631 18 AAW31884 HCV NS3A4 polyep

39 101 6.1 686 22 AAB62633 HCV antigen combin

40 101 6.1 1021 18 AAW34481 Fusion protein c20

41 101 6.1 1021 19 AAW40039 HCV NS35 polyep

42 101 6.1 1771 22 AAB62631 HCV delNS35 polyep

43 101 6.1 1771 22 AAB62632 Amino acid sequenc

44 101 6.1 1771 22 AAB62634 Amino acid sequenc

45 101 6.1 1771 22 AAB62635

ALIGNMENTS

RESULT 1

AA198464
ID AAY68464 standard; Protein; 299 AA.

XX AC AAY68464;

XX DT 25-APR-2000 (first entry)

XX DE Human testis specific factor tesmin SEQ ID NO:5.

XX KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;
KW differentiation regulatory factor; male germ cell regulatory actor;
KW germ cell differentiation; sterility.

XX OS Homo sapiens.

XX PN WO200004147-A1.

XX PD 27-JAN-2000.

XX PF 16-JUL-1999; 99WO-JP03859.

XX PR 17-JUL-1998; 98JP-0219856.

XX PA (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX PI Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

XX DR WPI: 2000-147785/13.

XX DR N-PSDB; AAZ88156.

XX PT New male germ cell regulatory factor tesmin expressed in spermatocytes
PT useful for investigation of germ cell differentiation and sterility -

XX Claim 1; Page 50-52; 63pp; Japanese.

XX The present sequence represents a male germ cell regulatory factor

CC expressed specifically in spermatocytes, designated tesmin.

CC CC can be used in the investigation of the mechanisms of germ cell

CC differentiation and sterility.

XX SQ Sequence 299 AA;

Query Match 100.0%; Score 1647; DB 21; Length 299;

Best Local Similarity 100.0%; Pred. No. 2.4e-141;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVICOLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYLOSDVPKPMALVGRFLPASTK 60

Db 1 mvicqlkggtqmlcidsrtrelkalhlvpqyqdnnylqsdvpkpmalvgrflpastk 60

Qy 61 LNLITQLEGALPSVWNGSAFPGSTLPGPKITLAGYCDCAFSGDFCNCCNCCNNL 120

Db 61 lnlitqlegalpsvngsafpgstlpgpkitlagycdcfafsgdfcnccnccnnl 120

Qy 121 HHDIERFKAIRKACLRNPEAFQPKIGKQGLGNVQPHNKGNCRRSGCLKNYCEYEAQI 180

Db 121 hhdierrfkaikacigrnpeafqpkigkqglgnvqphnkgncrrsgclknycceyeaqi 180

Qy 181 MCSSICKICGCKNYEESPERKTLMSMPNYMOTGLEGSHYLPPTKFSGLPRFSDHRRPSS 240

Db 181 mcssickicgcknyeesperktlmsmpnymotglegshylypptkfsglprfshdrpss 240

Qy 241 CISWEVVEATCACLQAQEEAEKHCSCLAEQMILEFGRCLSQIILHTEFKSKGLKME 299

Db 241 ciswevveatcacllaqeaeekhcscklaeqmileeefgrclsqilhteifskglkme 299

RESULT 2

AY68463

ID AAY68463 standard; Protein; 295 AA.

AC AAY68463;

DT 25-APR-2000 (first entry)

DE Mouse testis specific factor tesmin SEQ ID NO:4.

KW Testis specific factor; tesmin; cell death; regulation; spermatocyte;

KW differentiation regulatory factor; male germ cell regulatory actor;

KW germ cell differentiation; sterility.

OS Mus musculus.

PN WO200004147-A1.

XX 27-JAN-2000.

XX 16-JUL-1999; 99WO-JP03859.

XX 17-JUL-1998; 98JP-0219856.

XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX Sugihara T, Wadhwa R, Kaul SC, Mitsui Y;

XX WPI; 2000-147785/13.

XX N-PSDB; AAZ88155, AAZ88157.

XX New male germ cell regulatory factor tesmin expressed in spermatocytes

PT useful for investigation of germ cell differentiation and sterility -

XX Claim 1; Page 47-49; 63pp; Japanese.

CC The present sequence represents a male germ cell regulatory factor

CC expressed specifically in spermatocytes, designated tesmin.

CC CC can be used in the investigation of the mechanisms of germ cell

CC differentiation and sterility.

XX SQ Sequence 295 AA;

Query Match 76.0%; Score 1251; DB 21; Length 295;

Best Local Similarity 76.3%; Pred. No. 2.1e-105;

Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

Qy 1 MVICOLKGGTQMLCIDSRTRELKALHLVPOYQDQNNYLOSDVPKPMALVGRFLPASTK 60

Db 1 mvicqlkggtqmlcidsrtrelkalhlvpqyqdnnylqsdvpkpmaltlvgrllvpak 60

Qy 61 LNLITQLEGALPSVWNGSAFPGSTLPGPKITLAGYCDCAFSGDFCNCCNCCNNL 120

Db 61 lnlitqlegalpsvngsafpgstlpgpkitlagycdcfafsgdfcnscs----cnnl 116

Qy 121 HHDIERFKAIRKACLRNPEAFQPKIGKQGLGNVQPHNKGNCRRSGCLKNYCEYEAQI 180

Db 117 rhelerfkaikacigrnpeafqpkmgkgrlgaaaklrhskgcnckrsgclknycceyaki 176

Qy 181 MCSSICKICGCKNYEESPERKTLMSMPNYMOTGLEGSHYLPPTKFSGLPRFSDHRRPSS 240

Db 177 mcssickicacknyeesperkmlmstphymepqdfesshyllspakfsgppklrknraqfs 236

Qy 241 CISWEVVEATCACLQAQEEAEKHCSCLAEQMILEFGRCLSQIILHTEFKSKGLKME 299

Db 237 ciswevveatcacllaqeaeekhcscpslaeqmileeefgrclsqilhieftkskglkie 295

RESULT 3

ABB62035

ID ABB62035 standard; Protein; 950 AA.

XX ABB62035;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 12897.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL06138.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Disclosure; SEQ ID NO 12897; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 03-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PR Rosen CA, Barash SC, Ruben SM;
 PR WPI: 2001-465570/50.
 PR N-PSDB; AAL01300.
 PR Isolated nucleic acid molecule encoding a reproductive system antigen
 PR is used in preventing, treating or ameliorating a medical condition -
 PR Claim 11; SEQ ID NO 3988; 1297pp + Sequence Listing; English.
 PR The present invention provides the protein and coding sequences of a
 PR number of human reproductive system related antigens. These can be used
 PR in the prevention and treatment of reproductive system disorders,
 PR including cancer. The present sequence is a protein of the invention.
 PR Sequence 147 AA;
 SQ

Query Match 27.1%; Score 446; DB 22; Length 147;
 Best Local Similarity 98.9%; Pred. No. 1.1e-32;
 Matches 87; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MVTICQLKGGTQMLCINSTRRELKALHLVPOYQDQNNYLOSDVPKPMFTALVGRFLPASTK 60
 Db 60 mvicqlkggtqmlcidsnrtrelkalhlvpqyqdgqnylqsdvvpkpmftalvgrflpastk 119
 QY 61 LNLITQLEGALPSVVGSAFPGSGTLP 88
 Db 120 lnltqlegalpssvvggsafpgsgtexp 147
 RESULT 5
 AAW83392
 ID AAW83392 standard; Protein; 438 AA.
 XX
 AC AAW83392;
 XX
 DT 29-MAR-1999 (first entry)
 XX
 DE Caenorhabditis elegans synMuv protein LIN-54.
 XX
 KW LIN-54; synthetic multivulvar; SynMuv; signal transduction;
 KW animal model; tumour suppressor; retinoblastoma; cancer; cancer;
 KW cell proliferation; gene therapy.
 XX
 OS Caenorhabditis elegans.
 XX
 PN WO9854299-A1.
 XX
 PD 03-DEC-1998.
 XX
 PF 28-MAY-1998; 98WO-US11043.
 XX
 PR 28-MAY-1997; 97US-0047996.
 XX
 PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
 XX
 PI Ceol C, Horvitz HR, Lu X;
 XX
 DR WPI: 1999-045362/04.
 DR N-PSDB; AAV72865.
 XX
 Novel LIN-37, -35, -55, -52, -53 and -54 gene from C. elegans -
 PT useful for treating diseases associated with altered levels of cell
 PT proliferation, e.g. carcinomas
 XX
 PS Claim 7; Fig 13; 70pp; English.
 XX
 CC This is the amino acid sequence of LIN-54, a novel protein of
 CC Caenorhabditis elegans. The lin-54 gene (see AAV72865) is a novel
 CC synthetic multivulvar (synMuv) gene involved in cell fate and cell
 CC proliferation, and is part of a pathway that may be used as a
 CC genetic and biochemical model system for tumour suppression and
 CC cancer in mammals. SynMuv pathway genes and proteins may be used
 CC to identify genes which are part of the mammalian pathway and
 CC to modulate this pathway. Pure nucleic acids (see AAV72859-65) encoding
 CC C. elegans SynMuv polypeptides selected from LIN-37, -35, -55, -52,
 CC -53, -54 and E2F-1 (see AAW83386-92), are new. Also claimed are:
 CC (1) nucleic acids (see AAV72866-67) encoding the mouse and human
 CC homologues of C. elegans LIN-54; (2) vectors containing the nucleic
 CC acids; (3) transgenic cells; (4) a pure mammalian SynMuv
 CC polypeptide; and (5) an antibody which binds to a SynMuv family
 CC protein. The SynMuv nucleic acids and polypeptides can be used to
 CC diagnose and treat, especially by gene therapy, conditions
 CC involving altered levels of cell proliferation, e.g. SynMuv-
 CC associated carcinomas.
 XX
 CC Sequence 438 AA;
 SQ

CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB7789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX
SQ Sequence 53 AA;

Query Match 16.9%; Score 278; DB 22; Length 53;
Best Local Similarity 100.0%; Pred. No. 5.5e-18;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 VEATCACLQAQGEAEKEHCKSLAQMILEEFGRCLSQILHTEFKSKGLKME 299
|||||
DB 1 veatcaccllaqgeaekehckslaegmilleefgrclsqilhtefkskglkme 53

RESULT 8

AAW93348
ID AAW93348 standard; Protein: 251 AA.

XX
AC AAW93348;

XX
DT 06-NOV-2001 (first entry)

XX
DE Human polypeptide, SEQ ID NO: 2895.

XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.

XX
OS Homo sapiens.

XX
PN EP130094-A2.

XX
PD 05-SEP-2001.

XX
PF 07-JUL-2000; 2000EP-0114089.

XX
PR 08-JUL-1999; 99JP-0194486.

XX
PR 11-JAN-2000; 2000JP-0118774.

XX
PR 02-MAY-2000; 2000JP-0183765.

XX
PA (HELI-) HELIX RES INST.

XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX
DR WPI: 2001-524255/58.

XX
DR N-PSDB; AAK94268.

XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

XX
PS Claim 8; SEQ ID NO 2895; 1380pp + sequence listing; English.

XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide

CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 251 AA;

Query Match 13.1%; Score 216; DB 22; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.6e-11;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVTCQLKGGTQMCLCIDSRTRELKALHLVPPQYODQNNYLOS 41
|||||
DB 210 mvicqlkgtqmcidnsrtrelkalhlvppqyodqnnylqs 250

RESULT 9

ABB68888

ID ABB68888 standard; Protein: 243 AA.

XX
AC ABB68888;

DT 26-MAR-2002 (first entry)

XX
DE Drosophila melanogaster polypeptide SEQ ID NO 33456.

XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX
OS Drosophila melanogaster.

XX
PN WO200171042-A2.

XX
PD 27-SEP-2001.

XX
PF 23-MAR-2001; 2001WO-US09231.

XX
PR 23-MAR-2000; 2000US-191637P.

XX
PR 11-JUL-2000; 2000US-0614150.

XX
PA (PEKE) PE CORP NY.

XX
PI Venter JC, Adams M, Li PWD, Myers EW;

XX
PI WPI: 2001-656860/75.

XX
DR N-PSDB; ABL12991.

XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX
PS Disclosure; SEQ ID NO 33456; 21pp + Sequence Listing; English.

XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBV2072).

XX
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 243 AA;

Query Match 12.1%; Score 200; DB 22; Length 243;
Best Local Similarity 25.5%; Pred. No. 4.4e-10;
Matches 59; Conservative 25; Mismatches 63; Indels 84; Gaps 7;

QY 136 RNPEAFOPKIGKGO-LGNVKPQHNKGNCRSRGCKLKNYCEYEAQIMCSCICKIGCKNY 194

Db 7 rsavdkadgkkgaggv-----kxcccrgsqiknydcygsmaictkfcrcvgrnt 61
QY 195 E-----ESPER-----KTL 203
Db 62 evrelvdpsvaksavkrqkaamsakaaakagidvgkalqvaastlalpgkal 121
QY 204 MSMPNYMOTGG---LEGSHYLP-----PTKFSGLPRFS-----HDR 236
Db 122 mppkytlvagkppmasshinpipsrpiataatparavkqaeppmpvnllipvrhddr 181
QY 237 RPSSCISWEVATCATCLAQGEAEKHCSCKCLABQMILEFFGRCLSOIL 287
Db 182 rdrnifvqpnaallecmilgateaeqlglnelqvclvleefmrgyknll 232

RESULT 10
ABB61657
ID ABB61657 standard; Protein: 1981 AA.
XX
AC ABB61657;
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 11763.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL05760.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 11763; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1981 AA;

Query Match 6.9%; Score 113; DB 22; Length 1981;
Best Local Similarity 21.3%; Pred. No. 0.47; Indels 58; Gaps 7;
Matches 44; Conservative 26; Mismatches 79; Indels 58; Gaps 7;

QY 44 PKPMTALVGRFLPASTKLNLTQQLLEGALPSVVGSAFPPSGSTLPQPPKITLAGYDCDCEFA 103

Db 356 pnttsilktkypss---nsntrnsnmpsnntsnssstetdssl----- 396
QY 104 SGDFCNMCN-----NCCNNLHHDIERKAIKACLRNPEAFQPKIGKQLCGNVXPOHN 158
Db 397 -----cscnptfrfntldfrsnannrtytklnlclsnssqsis---snsqcsnslcn 446
QY 159 KGCNCRSRCKLNYCYEAQIMC-----SSICKICIGCKNYEES----- 197
Db 447 stcnssnslcnsssssklcsnsgysssksklcsnsgysssksklpsnpslyfa 506
QY 198 PERKTLMSMP---NYMOTGLEGSHYL 221
Db 507 pgtkliesivaeinvnktgtgfenshkl 533

RESULT 11
AAW81840
ID AAW81840 standard; Protein: 699 AA.
XX
AC AAW81840;
DT 05-FEB-1999 (first entry)
XX
DE Human LFA-1 beta chain protein.
XX
KW LFA-1; beta chain; human; anti-inflammatory; soluble.
XX
OS Homo sapiens.
XX
PN JP10295375-A.
XX
PD 10-NOV-1998.
XX
PF 25-APR-1997; 97JP-0109752.
XX
PR 25-APR-1997; 97JP-0109752.
XX
PA (DAUC) DAIICHI PHARM CO LTD.
XX
DR WPI; 1999-038275/04.
DR N-PSDB; AAV64856.
XX
PT Nucleic acid encoding soluble LFA-1 protein - used to identify
PT anti-inflammatory agents for use with LFA-1 mediated activity
XX
PS Disclosure; Page 13; 14pp; Japanese.
XX
CC This sequence represents a soluble LFA-1 beta chain protein. The
CC encoding nucleic acids can be used for the evaluation of an
CC anti-inflammatory agent by a test for the combination between
CC ICAM protein and a soluble LFA-1 hetero dimer protein. The soluble
CC LFA-1 protein does not associate in the absence of a surfactant.
XX
SQ Sequence 699 AA;

Query Match 6.6%; Score 109.5; DB 20; Length 699;
Best Local Similarity 22.3%; Pred. No. 0.27;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVICLKGCTQMLCINDSRTRELKALHLPQYQDQNNYLQSDVPKPM TALVGRFLPASTK 60
Db 436 itvtvlpqcecrdrqsrsl--ch-----gkgflecicrcdctgygkncecqtg 486
QY 61 LNLITQQLLEGA-----LPSVVGSAFPPSGSTLPQPPKITLAGYDCDCEFA----- 103
Db 487 -grssqegscrcrkdnnslscglgdcvcgqclchtsdvpq--kliygygcecdtincer 543
QY 104 -SGDFCN-----CNCNCCNNLHHDIERF-----KAIKACLRNPEAFQPKIGKQLCGN 152
Db 544 yngqvccgpgrglcfcgkc--rchpgefacsqcqcertegcl--nprrvacs-grgr--- 595

Qy 153 VKPOHKGCCNRRSGCLKNYCEVYEAQIM-----CSSIC-----KCIGCKNYEESP 198
 Db 596 -----CRC-----NVCECHSGYQLPLCQCPGCPSPGKYISCAECLKEKGP 638

RESULT 12
 AAP80836
 ID AAP80836 standard; protein; 769 AA.
 AC AAP80836;
 XX
 DT 18-OCT-1990 (first entry)
 XX
 DE Beta subunit of human glycoproteins LFA-1, Mac-1 or p150,95
 DE encoded by cDNA clone 18.1.1.
 XX
 KW Immunoassays; autoimmune diseases; allograft rejection;
 KW graft-versus-host disease; leukocyte adhesion deficiency; cDNA clone;
 KW beta-subunit of human glycoproteins LFA-1, Mac-1 or p150,95.
 XX
 OS Homo sapiens.

PH Key Location/Qualifiers
 FT Modified-site 50
 FT Modified-site /note="Potential N-glycosylation site"
 FT Modified-site 116
 FT Modified-site /note="As above"
 FT Modified-site 212
 FT Modified-site /note="As above"
 FT Modified-site 244
 FT Modified-site /note="As above"
 FT Modified-site 501
 FT Modified-site /note="As above"
 FT Modified-site 642
 FT Modified-site /note="As above"
 FT Region 197..205
 FT Region /label=TRYPTIC PEPTIDE L-56a
 FT Region 258..269
 FT Region /label=TRYPTIC PEPTIDE M-58
 FT Region 311..324
 FT Region /label=TRYPTIC PEPTIDE P-61
 FT Region 340..351
 FT Region /label=TRYPTIC PEPTIDE M-52
 FT Region 367..375
 FT Region /label=TRYPTIC PEPTIDE L-65
 FT Region 472..479
 FT Region /label=TRYPTIC PEPTIDE P-20
 FT Region 489..498
 FT Region /label=TRYPTIC PEPTIDE P-18
 FT Region 528..539
 FT Region /label=TRYPTIC PEPTIDE L-56b

PN W08806592-A.
 XX
 XX 07-SEP-1988.
 XX
 XX 26-FEB-1988; 88WO-US00611.
 XX
 XX 26-FEB-1987; 87US-0019440.
 XX
 XX (DANA-) DANA FARBER CANCER.
 XX
 XX Springer T, Kishimoto T, Roberts T;
 DR WPI; 1988-271137/38.
 DR N-PSDB; AAN80863.
 XX
 XX Recombinant beta-sub-unit of human glycoprotein -
 XX used for treatment and diagnosis of diseases involving cellular
 XX adhesion and for producing antibodies
 XX
 XX ; Fig 1; 22pp; English.
 XX

CC The beta-subunit of the glycoproteins was isolated and the AA sequence of
 CC at least a portion of it detd. A synthetic oligo probe corresp. to the AA
 CC sequence was synthesised and used as probe for a genomic or cDNA library
 CC contg. a DNA sequence encoding the beta-subunit. The p150,95 protein was
 CC purified from hairy cell leukaemia spleens. Mac-1 was purified from
 CC leukocytes and LFA-1 was purified from the SKW3 T cell line. The cDNA can
 CC be used to produce recombinant beta-subunit in large amts. Monoclonal
 CC antibodies raised against the recombinant beta-subunits can be used in
 CC immunoassays to monitor diseases such as AIDS, autoimmune disease,
 CC allograft rejection, graft-versus-host disease and leukocyte adhesion
 CC deficiency (LAD). The beta-subunit cDNA clone can be used in prenatal
 CC diagnosis of LAD. Diagnosis of LAD uses a 1.8Kb EcoRI fragment encoding
 XX the beta-subunit of human LFA-1 as a probe.
 XX

SQ Sequence 769 AA;

Query Match 6.6%; Score 109.5; DB 9; Length 769;
 Best Local Similarity 22.3%; Pred. No. 0.3;
 Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

Qy 1 MVICQLKGGTQMLCIDNSRTRELKALHLVPOYQDQNNYVLOSDFKPMATLVGRFLPASTK 60
 Db 436 ivtvqlpqcecrdrsl--ch-----gkfglegicrdtyigkncecqtq 486
 Qy 61 LNLITQOLEGA-----LPSVVGSAFSPGSTLPGPKITLAGYDCDFA----- 103
 Db 487 -grsselegscrkdnnsicsglgdcvcgqclchtsdvpq--klyggycecdtincer 543
 Qy 104 -SGDFCNN-----CNCNNCCNNLHDIERF-----KAIKACLGRRNPEAFQPKIGKGQLGN 152
 Db 544 yngqvcgpggrglcfcgkc--rchpgfegsacqcerttegcl--nprrvces-grgr--- 595
 Qy 153 VKPOHKGCCNRRSGCLKNYCEVYEAQIM-----CSSIC-----KCIGCKNYEESP 198
 Db 596 -----CRC-----NVCECHSGYQLPLCQCPGCPSPGKYISCAECLKEKGP 638

RESULT 13
 AAR07113
 ID AAR07113 standard; protein; 769 AA.
 XX
 AC AAR07113;
 XX
 DT 24-JAN-1991 (first entry)
 XX
 DE Recombinant beta-subunit of LFA-1.
 XX
 KW Rhinovirus; LFA-1; Mac-1; p150,95; ICAM-1; autoimmune diseases;
 KW graft rejection.
 XX
 XX Key Location/Qualifiers
 FT Peptide 197..205
 FT /label=tryptic peptide L-56a
 FT Peptide 258..269
 FT /label=tryptic peptide H58
 FT Peptide 311..324
 FT /label=tryptic peptide P-61
 FT Peptide 340..351
 FT /label=tryptic peptide H-52
 FT Peptide 367..375
 FT /label=tryptic peptide L-65
 FT Peptide 472..479
 FT /label=tryptic peptide P-20
 FT Peptide 489..498
 FT /label=tryptic peptide P-18
 FT Peptide 528..539
 FT /label=tryptic peptide L-56b

PN W09010652-A.
 XX
 XX 20-SEP-1990.
 XX

Fri Jul 12 08:35:30 2002

KW Leukocyte function associated antigen; LFA-1; I domain; inhibitor;
KW ligand; binding; pathology; treatment; screening; detection;
KW inflammatory disease; autoimmune disease; reperfusion injury;
KW myocardial infarction; stroke; organ transplant;
KW respiratory distress syndrome; glomerulonephritis; Crohn's disease;
KW T cell mediated disease; rheumatoid arthritis; osteoarthritis;
KW spondylitis; thyroid associated ophthalmopathy; Behcet disease;
KW sepsis; asthma; chronic bronchitis; silicosis; pulmonary sarcoidosis;
KW fibrosis; cystic fibrosis; keloid formation; scar formation;
KW atherosclerosis; transplant rejection; lupus;
KW inflammatory bowel disease; ulcerative colitis; leukaemia;
KW atopic dermatitis; psoriasis; urticaria; uveitis.

XX Homo sapiens.

XX OS

XX WO200060355-A2.

XX 12-OCT-2000.

XX 03-APR-2000; 2000WO-US08841.

XX 02-APR-1999; 99US-0285477.

XX (ICOS-) ICOS CORP.

XX (ABBO) ABBOTT LAB.

XX Staunton D, Van Der Vieren M, Huth J, Fowler K, Orme M;

XX Olejniczak ET;

XX WPI; 2000-665028/64.

XX N-PSDB; AAA54361.

XX Identifying inhibitor of leukocyte function-associated antigen binding

XX to natural ligand for ameliorating inflammatory diseases, by measuring

XX antigen-ligand binding in presence and absence of test compound

XX Example 4; Page 63-65; 66pp; English.

XX Inhibitors of leukocyte function-associated antigen binding are
XX useful for manufacturing a medicament for ameliorating a pathology
XX arising from LFA-1 binding to its ligand such as inflammatory
XX diseases, autoimmune diseases, reperfusion injury, myocardial
XX infarction, stroke, organ transplant, adult respiratory distress
XX syndrome, acute glomerulonephritis, Crohn's disease, T cell mediated
XX diseases, rheumatoid arthritis, osteoarthritis, spondylitis, thyroid
XX associated ophthalmopathy, Behcet disease, sepsis, asthma, chronic
XX bronchitis, allergic respiratory distress syndrome, chronic pulmonary
XX inflammatory disease, such as chronic obstructive pulmonary disease,
XX silicosis, pulmonary sarcoidosis, fibrosis, cystic fibrosis, keloid
XX formation, scar formation, atherosclerosis, transplant rejection
XX disorders, such as graft versus host reaction and allograft
XX rejection, chronic glomerulonephritis, lupus, inflammatory bowel
XX disease, such as ulcerative colitis, proliferative lymphocyte
XX diseases, such as leukemia, and inflammatory dermatoses, such as
XX atopic dermatitis, psoriasis, urticaria, and uveitis.

XX Sequence 769 AA;

Query Match 6.6%; Score 109.5; DB 21; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.3;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVICQLKGTOMLCIDNSRTELKALHLVPOYQDQNNYLOSDVPKPMATLVGRELPASTK 60
Db 436 ivtvqlpqccrcdrsl--ch-----gkfglegicrctdyigkncectq 486
QY 61 LNLITQOLEGA-----LPSVVNGSAFPGSTLPQPPKITLAGYCDCEFA----- 103
Db 487 -grsqlegrckrdnnslicsglgdcvcgclchtsdvp--kliyqgcecdtincer 543
QY 104 -SGDFCNN-----CNCNCCNNLHHDIERF-----KAIKAGIGRNPEAFQPKIGKGOLGN 152

Db 544 yngqvcgpggrglcfcgkc--rchpgfegsaqcqcerttegcl--nprrrvecs-grgr--- 595
QY 153 VKPQHNKGCNCRRRSGCLKNYCEYEAQIM-----CSSIC-----KCIGCKNVEESP 198
Db 596 -----crc-----nvcechsgyqlplcqcpgpcpgkyiscaecikfekgp 638

Search completed: July 11, 2002, 08:17:01
Job time: 205 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:15:20 ; Search time 54.3 seconds
(without alignments)
529.111 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSQLHTEFKSKLKME 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	437	26.5	429	2 T23152	hypothetical prote
2	376	22.8	571	2 B84585	hypothetical prote
3	357	21.7	603	2 T08955	hypothetical prote
4	245.5	14.9	658	2 F71410	hypothetical prote
5	117	7.1	497	2 T27827	hypothetical prote
6	112	6.8	1737	2 T00209	MEGF8 protein - hu
7	109.5	6.5	769	1 IJHULM	leukocyte adhesion
8	106.5	6.5	3635	2 T10053	laminin alpha 5 ch
9	105.5	6.4	2180	2 T29764	hypothetical prote
10	104	6.3	775	2 T21436	hypothetical prote
11	103.5	6.3	1786	1 MMFSB1	laminin beta-1 cha
12	102	6.2	1790	1 MMFSB1	laminin beta-1 cha
13	101	6.1	1700	2 S08167	Balbani ring 3 pr
14	101	6.1	1819	2 A71928	cag island protein
15	101	6.1	3011	1 GNMV3	genome polyprotein
16	100.5	6.1	633	2 T24898	hypothetical prote
17	100	6.1	395	2 S44454	transcription fact
18	100	6.1	520	2 G84510	probable receptor-
19	99.5	6.0	677	2 C42125	trophozoite cystei
20	99	6.0	373	2 T47488	hypothetical prote
21	99	6.0	565	2 T16408	hypothetical prote
22	99	6.0	2195	2 T34264	hypothetical prote
23	98.5	6.0	574	2 B88465	protein B0244.8 [i
24	98.5	6.0	962	2 JC5571	subtilisin-like pr
25	98.5	6.0	3106	1 S53868	laminin alpha-2 ch
26	98.5	6.0	5376	2 T42215	zonadhesin - mouse
27	97.5	5.9	432	2 T37509	hypothetical prote
28	97.5	5.9	969	1 A39490	subtilisin-like pr
29	97.5	5.9	975	2 JC5570	subtilisin-like pr

30	97.5	5.9	1766	2 A42125	trophozoite cystei
31	97	5.9	382	2 T29339	hypothetical prote
32	97	5.9	798	2 A40526	integrin beta-7 ch
33	97	5.9	3010	1 A45573	genome polyprotein
34	96.5	5.9	290	2 G72858	AcOrf-70 protein -
35	96.5	5.9	1895	2 T15881	hypothetical prote
36	96.5	5.9	3759	2 A35085	trithorax protein
37	96	5.8	798	2 A28193	integrin beta-1 ch
38	96	5.8	3010	1 GNMV3	genome polyprotein
39	95.5	5.8	1801	1 MHRTS	laminin beta-2 cha
40	95	5.8	1927	2 G64585	cag pathogenicity
41	94	5.7	3011	1 GNMV3	genome polyprotein
42	93.5	5.7	772	2 S32659	integrin beta 2 ch
43	93.5	5.7	1599	2 T16210	hypothetical prote
44	93	5.6	1751	1 MMHUMH	laminin alpha-2 ch
45	93	5.6	1810	1 A32230	tenascin precursor

ALIGNMENTS

RESULT 1

T23152

hypothetical protein JC8.6b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T23152

R:Lightning, J.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19699

A:Accession: T23152

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-429 <WIL>

A:Cross-references: EMBL:Z82274; PIDN:CA805228.1; GSPDB:GN00022; CESP:JC8.6b

A:Experimental source: clone JC8

C:Genetics:

A:Gene: CESP:JC8.6b

A:Map position: 4

A:Introns: 42/3; 91/3; 125/1; 160/2; 289/3; 371/3

Query Match 26.5%; Score 437; DB 2; Length 429;
Best Local Similarity 41.3%; Pred. No. 6.8e-28;
Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;

Qy 98 YCDCFASGDFCNCCNCCNHLHDIERFKAICACLRNPEAFQPKIG--KGQLGNVKP 155

Db 182 YCDCFANGFCRDCNCKDCHNNIEYDSQRSRAIROSLERNPNAPFKIGIARGITDIER 241

Qy 156 OHNKGCCNRRSGCLKNYCEYEAQIMCSSICKICCKN-----YEES----PERKTL 203

Db 242 LHQKGCHCKKSGCLKNYCEYEAQVPCDRCKCKCQNTETRYMTRYKNSGGAVSNTNAL 301

Qy 204 MSMPNMYMT-----GGL-----EGSHYLPPTKFGSLPRFSDHRR--PSSCISWEVVEAT 250

Db 302 MSLTNASTATPDSGPGSVVTDEHGDY-EDMLLSHKPKVEMDPRFPWYMTDEVVEAA 360

Qy 251 CACLLAQGEA-----BKEHCCKCLAEQMILEFGRCLSQL 287

Db 361 TCMVVAQAEALNVEKVTQDEDEKLNMEKLVLRFGRCLEQMI 403

RESULT 2

B84585

hypothetical protein At2g20110 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: B84585

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999

```

Query Match      21.78; Score 357; DB 2; Length 603;
Best Local Similarity 28.2%; Pred. No. 2.9e-21;
Matches 96; Conservative 43; Mismatches 83; Indels 118; Gaps 12;

Qy 39 LQSDVP---KP-MTALVGRF-----LPASTKLNLIQTQLEGALPSVVGSAFFSGSTLP 88
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 58 ISSSVPTIRPGMTAIGQVTVQVRPTLPMTATMSPPSQSQ-----IVNA---PIRHP 109
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 89 GPKPI-----TLAGYDCFASGDFCNKNCNCCNCLNLIHH 122
      || || || || || || || || || || || || || || || || || || || || ||
Db 110 ESPKARGPRPNVEGRDGTPOKKKOCNKHRSCLKLYCFCAFSGTYCDKCNVCNCFNNVDN 169
      || || || || || || || || || || || || || || || || || || || || ||

```

RESULT 5
T27827
hypothetical protein ZK287.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27827
R:McMurray, A.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z20425
A:Accession: T27827

Qy 265 HC-----SKCLA 271
 ||
 Db 974 HCVWSSSLQQLCS 986

RESULT 7

IJHULM
 N:Leukocyte adhesion protein beta chain (CD18) precursor - human
 N:Alternate names: integrin beta-2; LFA-1 beta chain; Mac-1 beta chain; p150,95 beta
 C:Species: Homo sapiens (man)
 C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 23-Jul-1999
 C:Accession: A25967; A91084; S19324; I70090; I70091; A29265; A94497
 R:Kishimoto, T.K.; O'Connor, K.; Lee, A.; Roberts, T.M.; Springer, T.A.
 Cell 48, 681-690, 1987
 A:Title: Cloning of the beta subunit of the leukocyte adhesion proteins: homology to
 A:Reference number: A25967; MUID:87131080
 A:Accession: A25967
 A:Molecule type: mRNA
 A:Residues: 1-769 <IS>
 A:Cross-references: GB:M15395; NID:g186933; PIDN:AAA59490.1; PID:g307113
 R:Note: source of LFA-1 was the SKW3 T-cell line; source of Mac-1 was pooled leukocytes
 A:Law, S.K.A.; Gagnon, J.; Hildreth, J.E.; Wells, C.E.; Willis, A.C.; Wong, A.J.
 EMO J. 6, 915-919, 1987
 A:Title: The primary structure of the B-subunit of the cell surface adhesion glycoprotein
 A:Reference number: A91084; MUID:87246525
 A:Accession: A91084
 A:Molecule type: mRNA
 A:Residues: 9-198, 'P', 200-769 <LA2>
 A:Cross-references: GB:Y00057; NID:g30228; PIDN:CAA68266.1; PID:g762939
 R:Weitzman, J.B.; Wells, C.E.; Wright, A.H.; Clark, P.A.; Law, S.K.A.
 FEBS Lett. 294, 97-103, 1991
 A:Title: The gene organisation of the human beta2 integrin subunit (CD18).
 A:Reference number: S19324; MUID:92077153
 A:Accession: S19324
 A:Molecule type: DNA
 A:Residues: 1-3; 16-23; 46-53; 106-113; 163-170; 244-251; 296-303; 328-335; 358-405-412-4
 R:Nelson, C.; Rabb, H.; Arnaout, M.A.
 J. Biol. Chem. 267, 3351-3357, 1992
 A:Title: Genetic cause of leukocyte adhesion molecule deficiency. Abnormal splicing a
 A:Reference number: I55376; MUID:92147694
 A:Accession: I70090
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 347-350, 'S', 352-355 <NEL>
 A:Cross-references: GB:S81234; NID:g245294; PIDN:AAB21404.1; PID:g245295
 A:Note: mutant form
 A:Accession: I70091
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 246-247, 'PSSQ', 248-249 <RES>
 A:Cross-references: GB:S81252; NID:g245289; PIDN:AAB21402.1; PID:g245290
 A:Note: mutant splice form
 C:Comment: The leukocyte adhesion proteins are noncovalently linked heterodimers of d
 and a deficiency of them is attributed to a genetic defect in the expression or stru
 C:Comment: The cysteine residues are involved in intrachain disulfide bonds.
 C:Genetics:
 A:Gene: GDB:ITGB2
 A:Cross-references: GDB:120574; OMIM:600065
 A:Map position: 21q22.3-21q22.3
 C:Superfamily: integrin beta chain; laminin-type EGF-like homology
 C:Keywords: cell adhesion; cytoskeleton; duplication; glycoprotein; heterodimer; leuk
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-769/Product: leukocyte adhesion protein beta chain #status predicted <MAT>
 F:23-670/Domain: extracellular #status predicted <EXT>
 F:445-631/Region: cysteine-rich
 F:459-540/Region: duplication
 F:541-627/Region: duplication
 F:701-723/Domain: transmembrane #status predicted <MEM>
 F:724-769/Domain: intracellular #status predicted <CYT>
 F:23/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predic
 F:50,116,212,254,501/Binding site: carbohydrate (Asn) (covalent) #status predicted

Db 937 PVTQLACVCDPGYIGSRCDDCASGFFGNPDSFGSCQPC--QCHNIDT-----T 985
QY 137 NPEAFQPKTKG-----GOLGNVYKQHNKCNCRSGCLKNY-----C 173
Db 986 DPEACDKDGRCLKLYHTDEGHCQLCQYGYGDALRQDCRKCVCNLYLGTVKEHCNGSDC 1045
QY 174 EYEAQIMCSSTCKIGCKNYEESPERKTLMSMPNYMQ-----TG-----GLESHYLPPT- 224
Db 1046 HCDKATGQCSCLPNVIG-----QNCDR-----CAPNTWQLASGTGCGPCNCNAHAFSGPSC 1096
QY 225 -KFSG---LPRFSDRRPSSC--ISW-----EVEATCACLLAAQ 257
Db 1097 NEFTGQCCMPGFG-GRTSECOELFWGDDPVECRACDCDPRGIETPPQDQSTGQCVVE 1155
QY 258 GEEAEKEHCKSC 269
Db 1156 GVEGPR--CDKC 1165
RESULT 12
MMFFB1
N:Alternate names: laminin chain B1
C:Species: Drosophila melanogaster
C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 19-Jan-2001
C:Accession: A28783; S14462; B28783
R:Montell, D.J.; Goodman, C.S.
Cell 53, 463-473, 1988
A:Title: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals do
A:Reference number: A28783; MUID:88210471
A:Accession: A28783
A:Molecule type: mRNA
A:Residues: 1-1790 <MON1>
A:CROSS-references: EMBL:M19525
R:Montell, D.J.; Goodman, C.S.
submitted to the EMBL Data Library, June 1988
A:Description: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reve
A:Reference number: S14462
A:Accession: S14462
A:Molecule type: mRNA
A:Residues: 1-667, 'L', 669-725, 'V', 728-947, 950-1790 <MON2>
A:CROSS-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802
C:Genetics:
A:Gene: lamB1
A:CROSS-references: FlyBase:FBgn0002527
A:Map position: 2L 28D
C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
C:Function:
A:Description: Interact with cells and with other basement membrane proteins to promote
C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1790/Product: laminin beta-1 chain #status predicted <MAT>
F:27-288/Domain: VI <DOM6>
F:289-561/Domain: V <DOM5>
F:290-354/Domain: laminin-type EGF-like homology <LE01>
F:357-417/Domain: laminin-type EGF-like homology <LE02>
F:420-477/Domain: laminin-type EGF-like homology <LE03>
F:480-528/Domain: laminin-type EGF-like homology <LE04>
F:531-561/Domain: laminin-type EGF-like homology #status atypical <LE05>
F:562-789/Domain: IV <DOM4>
F:643-645/Region: cell attachment (R-G-D) motif
F:790-1189/Domain: III <DOM3>
F:791-836/Domain: laminin-type EGF-like homology <LE06>
F:839-882/Domain: laminin-type EGF-like homology <LE07>
F:885-932/Domain: laminin-type EGF-like homology <LE08>
F:935-990/Domain: laminin-type EGF-like homology <LE09>
F:968-972/Region: cell adhesion #status predicted
F:993-1042/Domain: laminin-type EGF-like homology <LE10>
F:1045-1093/Domain: laminin-type EGF-like homology <LE11>
F:1096-1141/Domain: laminin-type EGF-like homology <LE12>
F:1144-1188/Domain: laminin-type EGF-like homology <LE13>
F:1190-1407/Domain: II <DOM2>

F:1408-1434/Domain: alpha <ALP>
F:1435-1790/Domain: I <DOM1>
F:51-56/Disulfide bonds: #status predicted
F:140,203,234,489,593,1053,1248,1303,1332,1343,1475,1517,1583,1646,1705/Binding
F:1191,1194,1788/Disulfide bonds: interchain #status predicted

Query Match 6.2%; Score 102; DB 1; Length 1790;

Best Local Similarity 21.4%; Pred. No. 3.8;
Matches 49; Conservative 14; Mismatches 88; Indels 78; Gaps 11;

QY 97 GYCD-----CF--ASGDFCNCC-----NCCNC-----CNLHHDIEREKA 129
Db 1005 GNCDDRTGACLKCLYTTGDHCELCCKDGGFDALQNCQCCDFLGTNTTAHCDRTFG 1064

QY 130 IKACLGRLNPEAFQPKIGKQGLGNVYKQHNK-----GCNRRSGCLKNYCECEVAAQIM 181

Db 1065 QCPCL-----PNVGVRCDCQCAENHWKIASGEGCESCNDPIGALHEOCNSVTGQCQ 1116

QY 182 CS---SICKICIGCK-NYEESEPERKTLMSMPNYMQTGLGSHYLPPTKFS--GLPRFSDH 235

Db 1117 CKPFGGGRACNQCQAHYGNPNK-----COPCECDQGAADFQCD 1157

QY 236 RRPSSCISWEVVEATCACLLAQGEAEKEHCKSLAEQMILEEFGRCLS 284

Db 1158 RETGNCVCHGEGGYKNECARGYICQFPHCSPC-----GECFN 1196

RESULT 13
S08167

Balbani ring 3 protein - midge (Chironomus tentans)

C:Species: Chironomus tentans

C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 21-Jul-2000

C:Accession: S08167

R:Paulsson, G.; Lendahl, U.; Galli, J.; Ericsson, C.; Wieslander, L.

J. Mol. Biol. 211, 331-349, 1990

A:Title: The balbani ring 3 gene in Chironomus tentans has a diverged repetitive str

A:Reference number: S08167; MUID:90172404

A:Accession: S08167

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1700 <PAU>

A:CROSS-references: GB:X52263; NID:g7057; PIDN:CAA36506.1; PID:g7058

C:Genetics:

A:Gene: BR3

A:Map position: 4

C:Superfamily: unassigned Balbani ring proteins

Query Match 6.1%; Score 101; DB 2; Length 1700;

Best Local Similarity 22.6%; Pred. No. 4.3;
Matches 51; Conservative 20; Mismatches 87; Indels 68; Gaps 15;

QY 99 CDCFASGD-----FC--NNCNC-----NNCCNLLHHDIERFKAKACILGRNPEAF 141

Db 198 CSCEKGDGKCGSKTWCKNCRICPTAEAPAGCCSAPLKWDDD--KSCACPAKWEKK 255

QY 142 QPKIGKQGLGNVYKQHNKGCNRRSGCLKN-----YCEYE-----AQIMCSICK 187

Db 256 EKCVESGKIWN---PNTCEGCAQLNCPDNKANKETCQCEKEVKKCGGVFCKDSCS 312

QY 188 CIGCKNYEESPERKTLMSMPNYMQTGLGSHYLP---PTKFSGLPR----- 231

Db 313 CV-CPGDKD---KTCTAPQVY---DGVACSCSPVNMQKPADCGCPQKWKDEKCEC 365

QY 232 -FSHRRRPSSCISWEVVEATCA-----LQAQGEAEKEHCS--KCL 270

Db 366 PVKHDCKNGKV--WDETICQICPRDPAVCTACKERGECSECKCI 409

RESULT 14

A71928

cag island protein - Helicobacter pylori (strain J99)

Result No.	Score	Query %			ID	Description
		Match	Length	DB		
1	1647	100.0	299	1	MT15_HUMAN	Q9Y415 homo sapien
2	1251	76.0	295	1	MT15_MOUSE	Q9WTJ6 mus musculu
3	109.5	6.6	769	1	ITB2_HUMAN	P05107 homo sapien
4	106.5	6.5	3718	1	LMA5_MOUSE	Q61001 mus musculu
5	103.5	6.3	1786	1	LMB1_MOUSE	P02469 mus musculu
6	103	6.3	769	1	ITB2_PIG	P53714 sus scrofa
7	102	6.2	1790	1	LMB1_DROME	PL1046 drosophila
8	101.5	6.2	396	1	TFH2_MOUSE	Q9J1B4 mus musculu
9	101	6.1	1700	1	BAR3_CHITE	Q03376 chiromomus
10	101	6.1	3011	1	POLG_HCV1	P26664 h genome po
11	100.5	6.1	760	1	EZ_DROME	P42124 drosophila
12	100	6.1	3695	1	LMA5_HUMAN	Q15230 homo sapien
13	99	6.0	787	1	ITB6_MOUSE	Q9Z0T9 mus musculu
14	98.5	6.0	3106	1	LMA2_MOUSE	Q06075 mus musculu
15	98.5	6.0	5376	1	ZAN_MOUSE	O88799 mus musculu
16	97.5	5.9	328	1	C170_GIALA	P15799 giardia lam
17	97.5	5.9	969	1	PAC4_HUMAN	P29122 homo sapien
18	97	5.9	395	1	TFH2_HUMAN	Q13888 homo sapien
19	97	5.9	798	1	ITB7_HUMAN	P26010 homo sapien
20	97	5.9	3010	1	POLG_HCVJT	Q00269 h genome po
21	96.5	5.9	290	1	Y070_NPVAC	P41470 autographa
22	96.5	5.9	1895	1	YLK3_CAEEL	P41951 caenorhabdi
23	96.5	5.9	3726	1	TRX_DROME	P26659 drosophila
24	96	5.8	798	1	ITB1_XENLA	P12606 xenopus lae
25	96	5.8	1808	1	TENA_CHICK	P10039 gallus galli
26	96	5.8	3010	1	POLG_HCVBK	P26663 h genome po
27	95.5	5.8	747	1	E2H1_HUMAN	Q92800 homo sapien
28	95.5	5.8	1801	1	LMB2_RAT	P15800 rattus norv
29	94	5.7	3011	1	POLG_HCVH	P27958 h genome po
30	93	5.6	3110	1	LMA2_HUMAN	P24043 homo sapien
31	93	5.6	3224	1	RBP2_HUMAN	P49792 homo sapien
32	92.5	5.6	332	1	XYNA_CRYAL	P07529 cryptococcu
33	92.5	5.6	611	1	LEM2_CANFA	P33730 canis famil

QY 181 MCSICKICGCKNYEESPERRKTLMSMPNMYMOTGLEGSHYLPPTKFSGLPRFSDRRPSS 240
 Db 181 MCSICKICGCKNYEESPERRKTLMSMPNMYMOTGLEGSHYLPPTKFSGLPRFSDRRPSS 240

QY 241 CISWVEVATCACLLAQGEAEKHCSCLAEOIMILEEFGRCLSQILHTEFKSKGLKME 299
 Db 241 CISWVEVATCACLLAQGEAEKHCSCLAEOIMILEEFGRCLSQILHTEFKSKGLKME 299

RESULT 2
 ID ITB2_MOUSE STANDARD; PRT; 295 AA.
 AC Q9WTU6;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Testin (Metallothionein-like 5, testis-specific) (Testis-specific
 DE metallothionein-like protein).
 GN MTL5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=Testis;
 RX MEDLINE=99208669; PubMed=10191092;
 RA Sugihara T., Wadhwa R., Kaul S.C., Mitsui Y.;
 RT "A novel testis-specific metallothionein-like protein, tesmin, is an
 RT early marker of male germ cell differentiation.";
 RL Genomics 57:130-136(1999).
 CC -1- FUNCTION: MAY HAVE A ROLE IN SPERMATOGENESIS.
 CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN SPERMATOCYTES.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED AS EARLY AS DAY 8 THAT COINCIDES
 CC WITH THE ENTRY OF GERM CELLS INTO MEIOSIS AND THEN INCREASED
 CC PROGRESSIVELY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U77383; AAD24667.1; -
 CC EMBL; U67176; AAD24666.1; -
 CC GDB; MG1:1340029; MTL5.
 KW Spermatogenesis..
 FT DOMAIN 99 187 CYS-RICH.
 SQ SEQUENCE 295 AA; 32487 MW; 516B9E1E1778251C CRC64;

Query Match 76.0%; Score 1251; DB 1; Length 295;
 Best Local Similarity 76.3%; Pred. No. 8.6e-101;
 Matches 228; Conservative 26; Mismatches 41; Indels 4; Gaps 1;

QY 1 MVICQLAGGQMLCIDNSRTELKALHLVPOYDONNYLQSDVPKPM TALVGRFLPASTK 60
 Db 1 MVICQLAGGQMLCIDNSRTELKALHLVPOYDONNYLQSDVPKPM TALVGRFLPASTK 60

QY 61 LNLITQDLEGALPSVNGSAFFSGSTLPGPKITLAGYCDCAFSGDCFNCCNCCNNL 120
 Db 61 LNLITQDNGALPSAVNGAAPPSPALQGPPTKILSGYCDCFSSGDFCNCS- - -CNNL 116

QY 121 HDIERFKAIKACUGRNPFAQPKIGKQGLGNVQPKQKNCNRRSGCLNKYCEYQAI 180
 Db 117 RHELERFKAIKACUDRNPFAQPKMGKGRGLGAARLRSKGNCKRSCLNKYCEYQAI 176

QY 181 MCSICKICGCKNYEESPERRKTLMSMPNMYMOTGLEGSHYLPPTKFSGLPRFSDRRPSS 240
 Db 177 MCSICKICGCKNYEESPERRKTLMSMPNMYMOTGLEGSHYLPPTKFSGLPRFSDRRPSS 236

QY 241 CISWVEVATCACLLAQGEAEKHCSCLAEOIMILEEFGRCLSQILHTEFKSKGLKME 299
 Db 237 CISWVEVATCACLLAQGEAEKHCSCLAEOIMILEEFGRCLSQILHTEFKSKGLKIE 295

RESULT 3
 ID ITB2_HUMAN STANDARD; PRT; 769 AA.
 AC P05107; O16418;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Integrin beta-2 precursor (Cell surface adhesion glycoproteins IFA-
 DE 1/CR3/p150.95 beta-subunit) (CD18) (Complement receptor C3 beta-
 DE subunit).
 GN ITGB2 OR CD18.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87131080; PubMed=3028646;
 RA Kishimoto T.K., O'Connor K., Lee A., Roberts T.M., Springer T.A.;
 RT "Cloning of the beta subunit of the leukocyte adhesion proteins:
 RT homology to an extracellular matrix receptor defines a novel
 RT supergene family.";
 RL Cell 48:681-690(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92077153; PubMed=1683838;
 RA Weitzman J.B., Wells C.E., Wright A.H., Clark P.A., Law S.K.A.;
 RT "The gene organisation of the human beta 2 integrin subunit (CD18).";
 RL FEBS Lett. 294:97-103(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
 RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
 RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Antonarakis S.E.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [4]
 RP SEQUENCE OF 9-769 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Spleen;
 RX MEDLINE=87246525; PubMed=2954816;
 RA Law S.K.A., Gagnon J., Hildreth J.E., Wells C.E., Willis A.C.,
 RA Wong A.J.;
 RT "The primary structure of the beta-subunit of the cell surface
 RT adhesion glycoproteins IFA-1, CR3 and p150,95 and its relationship to
 RT the fibronectin receptor.";
 RL EMBO J. 6:915-919(1987).
 RN [5]
 RP SEQUENCE OF 347-355 FROM N.A., AND VARIANTS LAD1 SER-351 AND TRP-586.
 RX MEDLINE=92147694; PubMed=1346613;
 RA Nelson C., Rabb H., Arnaout M.A.;
 RT "Genetic cause of leukocyte adhesion molecule deficiency. Abnormal
 RT splicing and a missense mutation in a conserved region of CD18 impair
 RT cell surface expression of beta 2 integrins.";
 RL J. Biol. Chem. 267:3351-3357(1992).
 RN [6]
 RP VARIANTS LAD1 PRO-149 AND ARG-169.
 RX MEDLINE=90293696; PubMed=1694220;
 RA Wardlaw A.J., Hibbs M.L., Stacker S.A., Springer T.A.;

QY 104 -SGDFCNN-----CNCNCCNLLHDIERF-----KAIKACLGRLNPEAFQPKIGKQGLGN 152
 Db 544 YNQGVGGPGRGLFCGKC- RCHPGFEGSACQERTTEGCL- NPPRRVECS-GRGR--- 595
 QY 153 VRPQHNKGNCRRSGGLKNVCYCEYEAQIM-----CSSIC-----KCIQCKNYEESP 198
 Db 596 -----CRC-----NVCECHSGYQLPLQCECPSPGCKYISCAECLKFEKGP 638

RESULT 4
 LMA5_MOUSE STANDARD; PRT; 3718 AA.
 AC Q61001; Q9JH06;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Laminin alpha-5 chain precursor.
 GN LMA5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE OF 1-92 FROM N.A.
 RA Timpl R., Sasaki T.;
 RT "Completion of the N-terminal sequence of the murine Laminin alpha 5
 chain";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 84-3718 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
 RX MEDLINE=96081906; PubMed=7499364;
 RA Miner J.H., Lewis R.M., Sanes J.R.;
 RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
 expression in adult mouse tissues";
 RL J. Biol. Chem. 270:28523-28526(1995).
 RN [3]
 RP REVISIONS.
 RA Miner J.H., Lewis R.M., Sanes J.R.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
 ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
 MEMBRANES (MAJOR COMPONENT).
 CC -1- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
 KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
 AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 CC EMBL; AJ293593; CAB999255.1; -

DR EMBL; U37501; AAC53430.1; -
 DR HSP; P02468; ITLE.
 DR MGD; MGI:105382; Lama5.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001886; LamNT.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR001791; Laminin_G.
 DR Pfam; PF00052; laminin_B; 1.
 DR Pfam; PF00053; laminin_EGF; 19.
 DR Pfam; PF00054; laminin_G; 2.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR ProDom; PD002082; LamNT; 1.
 DR ProDom; PD003031; Laminin_B; 1.
 DR SMART; SM00180; EGF_Lam; 17.
 DR SMART; SM00001; EGF_Like; 2.
 DR SMART; SM00281; Lamb; 1.
 DR SMART; SM00282; LamG; 5.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 19.
 DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
 DR PROSITE; PS50025; LAM_G_DOMAIN; 5.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 40
 FT CHAIN 1 40
 FT DOMAIN 41 304 LAMININ ALPHA-5 CHAIN.
 FT DOMAIN 305 363 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 364 433 LAMININ EGF-LIKE 1.
 FT DOMAIN 434 479 LAMININ EGF-LIKE 2.
 FT DOMAIN 500 546 LAMININ EGF-LIKE 3.
 FT DOMAIN 547 592 LAMININ EGF-LIKE 4.
 FT DOMAIN 593 637 LAMININ EGF-LIKE 5.
 FT DOMAIN 638 682 LAMININ EGF-LIKE 6.
 FT DOMAIN 683 728 LAMININ EGF-LIKE 7.
 FT DOMAIN 729 781 LAMININ EGF-LIKE 8.
 FT DOMAIN 782 833 LAMININ EGF-LIKE 9.
 FT DOMAIN 834 855 LAMININ EGF-LIKE 10.
 FT DOMAIN 856 1442 LAMININ EGF-LIKE 11 (INCOMPLETE).
 FT DOMAIN 1443 1488 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 1489 1532 LAMININ EGF-LIKE 12.
 FT DOMAIN 1533 1581 LAMININ EGF-LIKE 13.
 FT DOMAIN 1582 1632 LAMININ EGF-LIKE 14.
 FT DOMAIN 1633 1642 LAMININ EGF-LIKE 15.
 FT DOMAIN 1643 1831 LAMININ EGF-LIKE 16 (N-TERMINAL).
 FT DOMAIN 1832 1864 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 FT DOMAIN 1865 1914 LAMININ EGF-LIKE 17 (C-TERMINAL).
 FT DOMAIN 1915 1970 LAMININ EGF-LIKE 18.
 FT DOMAIN 1971 2024 LAMININ EGF-LIKE 19.
 FT DOMAIN 2025 2071 LAMININ EGF-LIKE 20.
 FT DOMAIN 2072 2118 LAMININ EGF-LIKE 21.
 FT DOMAIN 2119 2168 LAMININ EGF-LIKE 22.
 FT DOMAIN 2169 2735 DOMAIN II AND I.
 FT DOMAIN 2736 2933 LAMININ G-LIKE 1.
 FT DOMAIN 2947 3119 LAMININ G-LIKE 2.
 FT DOMAIN 3128 3296 LAMININ G-LIKE 3.
 FT DOMAIN 3337 3511 LAMININ G-LIKE 4.
 FT DOMAIN 3518 3689 LAMININ G-LIKE 5.
 FT DOMAIN 2205 2257 COILED COIL (POTENTIAL).
 FT DOMAIN 2230 2464 COILED COIL (POTENTIAL).
 FT DOMAIN 2604 2621 COILED COIL (POTENTIAL).
 FT DOMAIN 2639 2705 COILED COIL (POTENTIAL).
 FT SITE 1723 1725 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1839 1841 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 305 314 BY SIMILARITY.
 FT DISULFID 307 327 BY SIMILARITY.
 FT DISULFID 329 338 BY SIMILARITY.
 FT DISULFID 341 361 BY SIMILARITY.
 FT DISULFID 364 373 BY SIMILARITY.
 FT DISULFID 366 398 BY SIMILARITY.
 FT DISULFID 401 410 BY SIMILARITY.

Integrin beta-2 precursor (Cell surface adhesion glycoproteins LFA-1/CR3/PL150,95 beta-subunit) (CD18) (Complement receptor C3 beta-subunit).

ITGB2 OR CD18.

Sus scrofa (Pig).

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.

NCBI_TaxID=9823;

[1]

SEQUENCE FROM N.A.

Lee J.K., Schook L.B., Rutherford M.S.; "Molecular cloning and characterization of the porcine CD18 leukocyte adhesion molecule.";

xenotransplantation 3:222-230(1996).

- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2, ICAM3 AND ICAM4. INTEGRINS ALPHA-M/BETA-2 AND ALPHA-X/BETA-2 ARE RECEPTORS FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT AND FOR FIBRINOGEN. INTEGRIN ALPHA-X/BETA 2 RECOGNIZES THE SEQUENCE G-P-R IN FIBRINOGEN ALPHA-CHAIN. INTEGRIN ALPHA-M/BETA-2 RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN. INTEGRIN ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FACTOR X. INTEGRIN ALPHA-D/BETA-2 IS A RECEPTOR FOR ICAM3 AND VCAM1.

- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-2 ASSOCIATES WITH EITHER ALPHA-L, ALPHA-M, ALPHA-X OR ALPHA-D.

- SUBCELLULAR LOCATION: Type I membrane protein.

- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.

- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U13941; AAB16868.1; -

HSSP; P00743; IAP0.

InterPro; IPR000561; EGF-like.

InterPro; IPR002369; Integrin_B.

InterPro; IPR001169; Integrin_beta_C.

InterPro; IPR003659; PSI.

InterPro; IPR002035; VWFA.

Pfam; PF00362; Integrin_B; 1.

PRINTS; PR01186; INTEGRINB.

PRODOM; PR001811; Integrin_B; 1.

SMART; SM00187; INB; 1.

SMART; SM00423; PSI; 1.

SMART; SM00327; VWFA; 1.

PROSITE; PS00243; INTEGRIN_BETA; 3.

PROSITE; PS00022; EGF_1; UNKNOWN_2.

PROSITE; PS01186; EGF_2; UNKNOWN_3.

Integrin; Cell adhesion; Transmembrane; Glycoprotein; Repeat; Signal.

FT SIGNAL 1 22

FT CHAIN 23 769

FT DOMAIN 23 700

FT DOMAIN 23 700

FT TRANSMEM 701 723

FT DOMAIN 724 769

FT DOMAIN 124 363

FT DOMAIN 449 617

FT REPEAT 449 496

FT REPEAT 497 540

FT REPEAT 541 581

FT REPEAT 582 617

FT DISULFID 25 447

FT DISULFID 33 43

FT DISULFID 36 73

FT DISULFID 46 62

FT DISULFID 191 198

FT DISULFID 246 286

FT DISULFID 386 400

FT DISULFID 420 662

FT DISULFID 445 449

FT DISULFID 459 470 BY SIMILARITY.

FT DISULFID 467 506 BY SIMILARITY.

FT DISULFID 472 481 BY SIMILARITY.

FT DISULFID 483 497 BY SIMILARITY.

FT DISULFID 512 517 BY SIMILARITY.

FT DISULFID 514 549 BY SIMILARITY.

FT DISULFID 519 534 BY SIMILARITY.

FT DISULFID 536 541 BY SIMILARITY.

FT DISULFID 557 562 BY SIMILARITY.

FT DISULFID 559 590 BY SIMILARITY.

FT DISULFID 564 573 BY SIMILARITY.

FT DISULFID 575 582 BY SIMILARITY.

FT DISULFID 596 601 BY SIMILARITY.

FT DISULFID 598 643 BY SIMILARITY.

FT DISULFID 603 612 BY SIMILARITY.

FT DISULFID 615 618 BY SIMILARITY.

FT DISULFID 622 631 BY SIMILARITY.

FT DISULFID 628 695 BY SIMILARITY.

FT DISULFID 647 670 BY SIMILARITY.

FT SITE 397 399 CELL ATTACHMENT SITE (POTENTIAL).

FT CARBOHYD 50 50 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 254 254 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 642 642 N-LINKED (GLCNAC. .) (POTENTIAL).

SQ SEQUENCE 769 AA; 84789 MW; FDD0606CEEB850449 CRC64;

Query Match 6.3%; Score 103; DB 1; Length 769;

Best Local Similarity 21.5%; Pred. No. 0.22;

Matches 59; Conservative 28; Mismatches 88; Indels 100; Gaps 16;

QY 14 CIDNSTRRLKALHLPVQYQDONNVLSQDVPKPMYALYGRFLPASTKLNLTQQLEGALP 73

Db 449 CGDSKERTLOG-----NKGSMCEGVCRDAGYIGKHCECQTQ-GRSQELEGSCR 498

QY 74 SVVNGSAFPSTLPGPPKITLAGYCDCFAS-----GDFCNMNCNCCNLLHHDE 125

Db 499 K-DNSSIICSGL-----GDCICGQCVCHTSDVPNKKIYGQFC-ECDNMNC-----E 542

QY 126 RFKAIKACLG-----RNPEAFQPKIGKGLGNVYKPOHNKGCNRRS--GCLK-- 170

Db 543 RPDG-QVCGEKGKGLFCSTCRCEGFE-----GSACQCLKSTQGCNLQ 586

QY 171 -----NYCEC---YE-----AQIMCSSICKICGCKNYEESPERKTLMSM 206

Db 587 GVECSGRGRCNVCNQCDFGYQPLCTPCPCQVPCARYAKCAECLKFTGPFKNCSAE 646

QY 207 PNYMOTGLEGSHYLPPTKFKSGLPRFSHRRPSSC 241

Db 647 C-----GTTKLLPSRMGS--KCNERDSEGC 670

RESULT 7

LMBL DROME STANDARD; PRT; 1790 AA.

AC P11046; Q26328; Q9XZT4; Q9VLW6;

DT 01-JUL-1989 (Rel. 11, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Laminin beta-1 chain precursor (Laminin B1 chain).

GN LAMB1 OR LAMB1 OR CG7123.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CANTON-S;

RX MEDLINE=94000382; PubMed=8397815;

RA' Gow C.-H., Chang H.-Y., Lih C.-J., Chang T.-W., Hui C.-F.;

RT "Analysis of the Drosophila gene for the laminin B1 chain.";


```
FT DISULFID 887 901 BY SIMILARITY.
FT DISULFID 904 913 BY SIMILARITY.
FT DISULFID 916 932 BY SIMILARITY.
FT DISULFID 935 951 BY SIMILARITY.
FT DISULFID 937 962 BY SIMILARITY.
FT DISULFID 964 973 BY SIMILARITY.
FT DISULFID 976 990 BY SIMILARITY.
FT DISULFID 993 1007 BY SIMILARITY.
FT DISULFID 1017 1026 BY SIMILARITY.
FT DISULFID 1029 1042 BY SIMILARITY.
FT DISULFID 1045 1059 BY SIMILARITY.
FT DISULFID 1047 1066 BY SIMILARITY.
FT DISULFID 1068 1077 BY SIMILARITY.
FT DISULFID 1080 1093 BY SIMILARITY.
FT DISULFID 1096 1108 BY SIMILARITY.
FT DISULFID 1098 1115 BY SIMILARITY.
FT DISULFID 1117 1126 BY SIMILARITY.
FT DISULFID 1129 1141 BY SIMILARITY.
FT DISULFID 1144 1156 BY SIMILARITY.
FT DISULFID 1146 1163 BY SIMILARITY.
FT DISULFID 1165 1174 BY SIMILARITY.
FT DISULFID 1177 1188 BY SIMILARITY.
FT DISULFID 1191 1191 INTERCHAIN (PROBABLE).
FT DISULFID 1194 1194 INTERCHAIN (PROBABLE).
FT DISULFID 1788 1788 INTERCHAIN (PROBABLE).
FT SITE 643 645 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 140 140 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 203 203 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 593 593 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1053 1053 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1248 1248 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1303 1303 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. .) (POTENTIAL).

Query Match 6.2%; Score 102; DB 1; Length 1790;
Best Local Similarity 21.4%; Pred. No. 0.67;
Matches 49; Conservative 14; Mismatches 88; Indels 78; Gaps 11;

Qy 97 GYCD-----CF--ASGDFCNK-----NCNC-----C>NNLHDIERFKA 129
Db 1005 GNCDRQTGACLKCLYQTTGDHCELCCKDGFDDALQNCQCECDPLGNTNTIAHCDRETF 1064
Qy 130 IKACILGRNPEAFQPKIGKGLGNVYKPOHNK-----GCNCRSGCLKNKYCEYEAQIM 181
Db 1065 QCPCD-----PNVQGVRCDCQCAENHWKIASGEGCESCNDPIGALHEQCNSYTGOCQ 1116
Qy 182 CS---SICKIGCK-NYBESPERKTLMSMPNYMOTGGLESHYLPPTKFS--GLPRFSGHD 235
Db 1117 CKPGFGRACNOCQAHYWGNFK-----CQPCEDQFGAADFQCD 1157
Qy 236 RPPSSCISNVEVATCCLLAGGEAEKHEKSCLAEOIMLEEFGRCLUS 284
Db 1158 RETGNCVCHGIGGYKCNCEARGYIGQPPHGSPC-----GECFN 1196
```

```
RESULT 8
TFH2_MOUSE
ID TFH2_MOUSE STANDARD; PRT; 396 AA.
AC Q9JIB4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE TFIIH basal transcription factor complex p44 subunit (Basic
DE transcription factor 2.44 kDa subunit) (BTF2-p44) (General
DE transcription factor IIH polypeptide 2).
GN GTF2H2 OR BTF2P44.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20414747; PubMed=10958627;
RA Endrizzi M.G., Hadinoto V., Growney J.D., Miller W., Dietrich W.F.;
RT "Genomic sequence analysis of the mouse Naip gene array.";
RL Genome Res. 10:1095-1102(2000).
CC -!- FUNCTION: Component of the core-TFIIH basal transcription factor
CC involved in nucleotide excision repair (NER) of DNA and, when
CC complexed to CAK, in RNA transcription by RNA polymerase II. The
CC N-terminus interacts with and regulates XPD whereas an intact C-
CC terminus is required for a successful escape of RNAP II from the
CC promoter (By similarity).
CC -!- SUBUNIT: One of the six subunits forming the core-TFIIH basal
CC transcription factor. Interacts with XPB, XPD, p62 and p34 (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 C4-TYPE ZINC FINGER.
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF242432; AAF82753.1; -.
CC MGD; MGI:1345669; Gtf2h2.
CC InterPro; IPR001841; Znf_ring.
CC SMART; SM00327; VWA; 1.
CC PROSITE; PS00518; 2F_RING_1; FALSE_NEG.
CC PROSITE; PS00089; 2F_RING_2; FALSE_NEG.
CC Transcription regulation; DNA repair; Nuclear protein; zinc-finger.
CC ZN_FING 292 309 C4-TYPE.
CC ZN_FING 346 383 RING-TYPE.
CC DOMAIN 242 245 POLY-SER.
CC SEQUENCE 396 AA; 44687 MW; 1BC57AFD464AA83F CRC64;
```

```
Query Match 6.2%; Score 101.5; DB 1; Length 396;
Best Local Similarity 21.8%; Pred. No. 0.15;
Matches 48; Conservative 27; Mismatches 72; Indels 73; Gaps 11;

Qy 2 VICOLKGTQMLCIDSRTRELKALHLPQYQDQNNYLQSDVPKPMATLVGRFLPASTKL 61
Db 210 VLARETGTYYHVLDETHYKELLAHHVSPPPASSSS-----ECSLRMGFFQHTIA 260
Qy 62 NLITQOLEGALPSVVNGSAFPSTLPQPKITLAGYCDCEFASGDFCNKNCNNC----- 116
Db 261 SLSDQD---AKPS-FSMAHLDDNNSTPG--LTLGGY-----FCQCRAKYCELPVE 305
Qy 117 -----C>NNLHH-----DIERFKAKACILGRNPEAFQPKIGKGLGN 152
Db 306 CKICGLTLVSAPHLARSYHFLPDLDAFOEISLEEKYGERFCYGCQ-----GE 352
Qy 153 VKPOHNKGCN-CRRSGCLKNKYCEYEAQIMCSCICKICGC 191
Db 353 LRQDHVYVCTVCQNVFCVD--CDVF----VHDSLHCPCGC 386
```

```
RESULT 9
BAR3_CHITE
ID BAR3_CHITE STANDARD; PRT; 1700 AA.
AC Q03376;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Balbiani RING protein 3 precursor.
GN BR3.
OS Chironomus tentans (Midge).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
```

OC Chironomidae; Chironomidae; Chironominae; Chironomus.
 OX NCBI_TaxID=7153;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX MEDLINE=90172404; PubMed=1689777;
 RA Paulsson G., Lendahl U., Galli J., Ericsson C., Wieslander L.;
 RT "The Balbiani ring 3 gene in Chironomus tentans has a diverged
 RL J. Mol. Biol. 211:331-349(1990).
 CC -!- FUNCTION: USED BY THE LARVAE TO CONSTRUCT A SUPRAMOLECULAR
 CC STRUCTURE, THE LARVAL TUBE. BALBIANI RING PROTEIN 3 COULD PLAY A
 CC ROLE AS A TRANSPORT PROTEIN THAT BINDS TO OTHER PROTEINS
 CC INTRACELLULARLY AND IN THE GLAND LUMEN IN ORDER TO PREVENT THESE
 CC FROM FORMING WATER-INSOLUBLE FIBERS TOO EARLY.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: SALIVARY GLAND.
 CC -!- DOMAIN: HAS 82 APPROXIMATE REPEATS OF CYS-X-CYS-X-CYS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X52263; CAA36506.1; -
 CC PIR: S08167; S08167.
 CC HSP: P35555; IAPJ.
 CC InterPro: IPR004153; CXXC.
 CC Pfam: PF03128; CXXC; 71.
 CC Repeat: Signal.
 FT SIGNAL 1 20 POTENTIAL
 FT CHAIN 21 1700 BALBIANI RING PROTEIN 3.
 FT SEQUENCE 1700 AA; 186145 MW; 3420282521B0815 CRC64;
 SQ
 Query Match 6.1%; Score 101; DB 1; Length 1700;
 Best Local Similarity 22.6%; Pred. No. 0.77;
 Matches 51; Conservative 20; Mismatches 87; Indels 68; Gaps 15;
 QY 99 CDCFASGD-----FC-NNCNC-----NNCCNHLHDIERFKAIKACLGSRNPEAF 141
 Db 198 CSCECKGDGCKGSKIWCKNNCRICPTABPAGCSAPLKWDD--KSCACAPAKMBEKK 255
 QY 142 QPKTIGKQGLGNVQPHNKGNCNRSGCLKN-----YCECYE-----AQIMCSICK 187
 Db 256 EKVESGKIWN---PNTCECGCAQLNCPDNKANKETCCCECKEVKCNCGGVFCKDSCS 312
 QY 188 CIGCKNTEESPERKTLMSPNYMTGLGSHYLP---PTKFSGLPR----- 231
 Db 313 CV-CPGDKD---KTCYAPQVY---DGVACSCPCPVNMQRPADGCPRPQKWKRECRCEC 365
 QY 232 -FSHRRSPSCISWEVVEATCAC-----LIAQGEAEKEHCS-KCL 270
 Db 366 PVKHDKNGKV--WDETICQICPDAPVCTAGKRCGESECKCI 409
 RESULT 10
 POLG_HCV1 STANDARD; PRT; 3011 AA.
 AC P26664;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-1991 (Rel. 40, Last annotation update)
 DE Genome polypeptide [Contains: Capsid protein C (Core protein) (P22);
 DE Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2
 DE (GP68) (GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21)
 DE (EC 3.4.22.1); Protease/helicase NS3 (P70) (Hepacivirus)
 DE (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein
 DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein
 DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].
 OS Hepatitis C virus (isolate 1) (HCV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OC Hepacivirus.
 OX NCBI_TaxID=11104;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX MEDLINE=91172826; PubMed=1848704;
 RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,
 RA Bradley D.W., Kuo G., Houghton M.;
 RT "Genetic organization and diversity of the hepatitis C virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polyprotein, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M62321; AAA45676.1; -
 CC PIR: A39166; GNWVC3.
 CC HSP: P27958; LHEI.
 CC MEROPS: S29.001; -
 CC MEROPS: U39.001; -
 CC InterPro: IPR001410; DEAD.
 CC InterPro: IPR002531; HCV_NS1.
 CC InterPro: IPR002518; HCV_NS2.
 CC InterPro: IPR004109; HCV_NS3.
 CC InterPro: IPR000745; HCV_NS4a.
 CC InterPro: IPR001490; HCV_NS4b.
 CC InterPro: IPR002868; HCV_NS5a.
 CC InterPro: IPR002166; HCV_RdRP.
 CC InterPro: IPR002522; HCV_capsid.
 CC InterPro: IPR002521; HCV_core.
 CC InterPro: IPR002519; HCV_env.
 CC InterPro: IPR001650; Helicase_C.
 CC Pfam: PF01543; HCV_capsid; 1.
 CC Pfam: PF01542; HCV_core; 1.
 CC Pfam: PF01539; HCV_env; 1.
 CC Pfam: PF01560; HCV_NS1; 1.
 CC Pfam: PF01538; HCV_NS2; 1.
 CC Pfam: PF02907; HCV_NS3; 1.
 CC Pfam: PF01006; HCV_NS4a; 1.
 CC Pfam: PF01001; HCV_NS4b; 1.
 CC Pfam: PF01506; HCV_NS5a; 1.
 CC Pfam: PF00998; HCV_RdRP; 1.
 CC Pfam: PF00271; helicase_C; 1.
 CC ProDom: PD186062; HCV_NS1; 1.
 CC SMART: SM00492; HELIC3; 1.
 CC PolyProtein: Glycoprotein; Transferase; RNA-directed RNA polymerase;
 CC Envelope protein; Helicase; ATP-binding;
 CC Core protein; Coat protein; Envelope protein; Hydrolase; Serine protease.
 CC Transmembrane; Nonstructural
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).
 FT

```
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT ACT_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT NP_BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3011 AA; 327197 MW; 65F8C9447FCE5AF9 CRC64;

Query Match 6.1%; Score 101; DB 1; Length 3011;
Best Local Similarity 24.8%; Pred. No. 1.4;
Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

QY 79 SAFP-SGSLPGPKITLAGYDCDFASGDCNNCNCNNL-----HHDIERFRAI 130
DB 1424 SVIPTSQVWVATDALMTGY-----TGFDSVIDCCTVDTVDFSLDPTFTITL 1478
QY 131 KACLRNPEAFQPKIGKQLGQ---NVKQHNKGCNRRSGCLN--YCEYEAQIMCSSI 185
DB 1479 ODAVSRTQR--RGRTRGKPGIYRVAPGE-----RPSGMFDDSVLCGYDA----- 1523
QY 186 KCICCKNVEESPERKTLMSNPNYQOTGGL-----EG-----SHYLPPTKF 226
DB 1524 ---GCANYELTP-AETVRLRAYNNTGLPVCQDHLFEWGVFTGLTHIDAHFLSQTK 1578
QY 227 SG--LPRFSDRRPSSCISWEVVEATCACLQAQGEAEKEHCKSL 270
DB 1579 SGENLP-----YLVAYQATVCARAQAAPPSPWDQWVKCL 1611

RESULT 11
EZ_DROME STANDARD; PRT; 760 AA.
ID E2_DROME
AC P42124;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Enhancer of zeste protein.
GN E(2).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94019309; PubMed=8413234;
RA Jones R.S., Gelbart W.M.;
RT "The Drosophila Polycomb-group gene Enhancer of zeste contains a
region with sequence similarity to trithorax.";
```

```
RL Mol. Cell. Biol. 13:6357-6366(1993).
[2]
RN CHARACTERIZATION.
RX MEDLINE=97164720; PubMed=9012527;
RA Carrington E.A., Jones R.S.;
RT "The Drosophila Enhancer of zeste gene encodes a chromosomal protein:
examination of wild-type and mutant protein distribution.";
RL Development 122:4073-4083(1996).
CC -!- FUNCTION: NEGATIVELY REGULATES THE SEGMENT IDENTITY GENES OF THE
ANTENNAPEDIA (ANT-C) AND BITHORAX (BX-C) GENE COMPLEXES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: CONTAINS 1 SET DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE EZ FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DB EMBL; U00180; AAC46462.1; .
DR Flybase; FBgn0000629; E(z).
DR InterPro; IPR001005; Myb_DNA_bind.
DR InterPro; IPR001214; SET.
DR Pfam; PF00856; SET; 1.
DR SMART; SM00395; SANT; 1.
DR SMART; SM00317; SET; 1.
DR PROSITE; PS0280; SET; 1.
DR Transcription regulation; Nuclear protein; DNA-binding; Repressor;
KW Developmental protein.
FT DOMAIN 505 510 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 538 619 CYS-RICH.
FT DOMAIN 632 743 SET.
SQ SEQUENCE 760 AA; 86965 MW; 4A28A8EAD7968AC4 CRC64;

Query Match 6.1%; Score 100.5; DB 1; Length 760;
Best Local Similarity 20.3%; Pred. No. 0.36;
Matches 36; Conservative 23; Mismatches 63; Indels 55; Gaps 8;

QY 99 CDCFASGDFCNN-CNCCNCCNLLHDIERFKAICACLRNPEAFQPKIGKGLGNVKKQH 157
DB 549 CSCIQTNQCFKFCNCSSDCQN-----RF----- 572
QY 158 NKGCCNRRSGCLKNKYCEYEAQIMCS-SICKCIGCKNTEESPERKTLMSNPNYQOTGGL 216
DB 573 -PGCRC-KAQCNTKQPCYLAVRECDPDLQACGADQF-----KLTKITCKNVQVQGLH 625
QY 217 GSHYLPPTKFGSLPRFSDRRPSSCISWEVVEATCACLQAQ-----GEEAEKEHCS 267
DB 626 KHLMLAPSDIAGWGIFLKEGAQKN-----EFISEYCEIISQDEADRRGRKVDKYMCS 678

RESULT 12
LMA5_HUMAN STANDARD; PRT; 3695 AA.
ID LMA5_HUMAN
AC O15230; Q9H1P1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LMA5 OR KIA0533 OR KIAA1907.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94019309; PubMed=8413234;
RA Jones R.S., Gelbart W.M.;
RT "The Drosophila Polycomb-group gene Enhancer of zeste contains a
region with sequence similarity to trithorax.";
```


RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clapp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leivasalho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaubin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.,
RL "The DNA sequence and comparative analysis of human chromosome 20." ;
RT Nature 414:865-871(2001).
RN [2]
RP SEQUENCE OF 197-1934 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21456161; PubMed=11572484;
RA Nagase T., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins." ;
RL DNA Res. 8:179-187(2001).
RN [3]
RP SEQUENCE OF 2051-3695 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98290545; PubMed=9629581;
RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro." ;
RL DNA Res. 5:31-39(1998).
RN [4]
RP SEQUENCE OF 2743-3695 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=97415425; PubMed=9271224;
RA Durkin M.E., Loebel F., Mattei M.-G., Gilpin B.J., Albrechtsen R.,
RA Wewer U.M.;
RT "Tissue-specific expression of the human laminin alpha5-chain, and
RT mapping of the gene to human chromosome 20q13.2-13.3 and to distal
RT mouse chromosome 2 near the locus for the ragged (Ra) mutation." ;
RL FEBS Lett. 411:296-300(1997).
RN [5]
RP EXPRESSION IN RETINA.
RX MEDLINE=20422761; PubMed=10964957;
RA Libby R.T., Champlaud M.-F., Claudepierre T., Xu Y., Gibbons E.P.,
RA Koch M., Burgeson R.E., Hunter D.D., Brunken W.J.;
RT "Laminin expression in adult and developing retinae: evidence of two
RT novel CNS laminins." ;
RL J. Neurosci. 20:6517-6528(2000).
CC -!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION AND ORGANIZATION
CC OF CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
CC -!- SUBUNIT: LAMININ-15 COMPLEX IS AN HETEROPOLYMER COMPOSED OF THREE
CC CHAINS (ALPHA-5/BETA-2/GAMMA-3) WHICH ARE BOUND TO EACH OTHER BY
CC DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE COMPRISING ONE LONG
CC AND THREE SHORT ARMS WITH GLOBULES AT EACH END.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN HEART, LUNG, KIDNEY, SKELETAL
CC MUSCLE, PANCREAS, RETINA AND PLACENTA. LITTLE OR NO EXPRESSION IN

CC BRAIN AND LIVER.
CC -!- DOMAIN: DOMAIN G IS GLOBULAR AND IS PART OF THE MAJOR CELL-BINDING
CC SITE LOCATED IN THE LONG ARM OF THE LAMININ HETEROPOLYMER.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL354836; CAC22309.1; ALT_SEQ.
CC EMBL: AB067494; BAB67800.1; -
CC EMBL: AB011105; BAA25459.1; -
CC EMBL: Z95636; CAB09137.1; -
CC HSSP: P02468; IKLO.
CC MIM: 601033; -
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR002049; Laminin_EGF.
CC InterPro: IPR001791; Laminin_G.
CC Pfam: PF00053; laminin_EGF; 1.
CC Pfam: PF00054; laminin_G; 2.
CC SMART: SM00180; EGF_Lam; 2.
CC SMART: SM00282; LamG; 5.
CC PROSITE: PS00022; EGF_1; 19.
CC PROSITE: PS01186; EGF_2; 3.
CC PROSITE: PS01248; LAMININ_TYPE_EGF; 16.
CC PROSITE: PS50025; LAM_G_DOMAIN; 5.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC SIGNAL 1 35 POTENTIAL.
CC FT CHAIN 36 3695 LAMININ ALPHA-5 CHAIN.
CC FT DOMAIN 36 299 LAMININ N-TERMINAL (DOMAIN VI).
CC FT DOMAIN 300 358 LAMININ EGF-LIKE 1.
CC FT DOMAIN 359 428 LAMININ EGF-LIKE 2.
CC FT DOMAIN 429 474 LAMININ EGF-LIKE 3.
CC FT DOMAIN 494 540 LAMININ EGF-LIKE 4.
CC FT DOMAIN 541 586 LAMININ EGF-LIKE 5.
CC FT DOMAIN 587 631 LAMININ EGF-LIKE 6.
CC FT DOMAIN 632 676 LAMININ EGF-LIKE 7.
CC FT DOMAIN 677 722 LAMININ EGF-LIKE 8.
CC FT DOMAIN 723 775 LAMININ EGF-LIKE 9.
CC FT DOMAIN 776 828 LAMININ EGF-LIKE 10.
CC FT DOMAIN 829 850 LAMININ EGF-LIKE 11 (INCOMPLETE).
CC FT DOMAIN 851 1437 LAMININ DOMAIN IV 1 (DOMAIN IV B).
CC FT DOMAIN 1438 1483 LAMININ EGF-LIKE 12.
CC FT DOMAIN 1484 1527 LAMININ EGF-LIKE 13.
CC FT DOMAIN 1528 1576 LAMININ EGF-LIKE 14.
CC FT DOMAIN 1577 1627 LAMININ EGF-LIKE 15.
CC FT DOMAIN 1628 1637 LAMININ EGF-LIKE 16 (N-TERMINAL).
CC FT DOMAIN 1638 1830 LAMININ DOMAIN IV 2 (DOMAIN IV A).
CC FT DOMAIN 1831 1863 LAMININ EGF-LIKE 17.
CC FT DOMAIN 1864 1912 LAMININ EGF-LIKE 18.
CC FT DOMAIN 1913 1968 LAMININ EGF-LIKE 19.
CC FT DOMAIN 1969 2022 LAMININ EGF-LIKE 20.
CC FT DOMAIN 2023 2069 LAMININ EGF-LIKE 21.
CC FT DOMAIN 2070 2116 LAMININ EGF-LIKE 22.
CC FT DOMAIN 2117 2166 LAMININ EGF-LIKE 23.
CC FT DOMAIN 2167 2735 DOMAIN II AND I.
CC FT DOMAIN 2736 2929 LAMININ G-LIKE 1.
CC FT DOMAIN 2941 3115 LAMININ G-LIKE 2.
CC FT DOMAIN 3124 3292 LAMININ G-LIKE 3.
CC FT DOMAIN 3340 3513 LAMININ G-LIKE 4.
CC FT DOMAIN 3520 3692 LAMININ G-LIKE 5.
CC FT DOMAIN 2203 2221 COILED COIL (POTENTIAL).
CC FT DOMAIN 2335 2466 COILED COIL (POTENTIAL).
CC FT DOMAIN 2510 2670 COILED COIL (POTENTIAL).
CC FT SITE 1722 1724 CELL ATTACHMENT SITE (POTENTIAL).

J. Immunol. 151:1789-1801(1993).

[3]

SEQUENCE OF 64-281 FROM N.A.
MEDLINE=9519718; PubMed=7874173;
Xu H., Wu X.R., Wewer U.M., Engvall E.;
"Murine muscular dystrophy caused by a mutation in the laminin alpha
2 (Lama2) gene.";
Nat. Genet. 8:297-302(1994).

[4]

X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.
MEDLINE=20085745; PubMed=10619025;
Hohenester E., Tisi D., Talts J.F., Timpl R.;
"The crystal structure of a laminin G-like module reveals the
molecular basis of alpha-dystroglycan binding to laminins, perlecan,
and agrin.";
Mol. Cell 4:783-792(1999).

-!- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT. BY INTERACTING
WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.

-!- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
THE ALPHA-2 CHAIN IS A SUBUNIT OF LAMININ-2 (MEROSIN) AND LAMININ-
4 (S-MEROSIN).

-!- SUBCELLULAR LOCATION: Extracellular.

-!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
COMPONENT).

-!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.

-!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.

-!- DISEASE: DEFECTS IN LAMA2 ARE A CAUSE OF MURINE MUSCULAR DYSTROPHY
(DY2J).

-!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).

-!- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.

-!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U12447; AAC52165.1; -;
EMBL; X69869; CAA49502.1; -;
EMBL; S75315; AAB33573.1; -;
PDB; 1Q0U; 03-DEC-99.
MGD; MGI:99912; Lama2,
InterPro; IPR000561; EGF-like.
InterPro; IPR001886; LamNT.
InterPro; IPR000034; Laminin_B.
InterPro; IPR002049; Laminin_EGF.
InterPro; IPR001791; Laminin_G.
Pfam; PF00052; laminin_B_2.
Pfam; PF00053; laminin_EGF; 15.
Pfam; PF00054; laminin_G; 5.
Pfam; PF00055; laminin_Nterm; 1.
PRINTS; P00011; EGFLAMININ.
PRODOM; PD002082; LamNT; 1.
ProDom; PD003031; Laminin_B; 2.
SMART; SM00180; EGF_Lam; 15.
SMART; SM00001; EGF_Like; 1.
SMART; SM00281; LamB; 2.
SMART; SM00282; LamG; 5.
SMART; SM00136; LamNT; 1.
PROSITE; PS00022; EGF_1; 11.
PROSITE; PS01186; EGF_2; 3.
PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
PROSITE; PS50025; LAM_G_DOMAIN; 5.

KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.

FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 3106 LAMININ ALPHA-2 CHAIN.
 FT DOMAIN 23 282 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 283 339 LAMININ EGF-LIKE 1.
 FT DOMAIN 340 409 LAMININ EGF-LIKE 2.
 FT DOMAIN 410 464 LAMININ EGF-LIKE 3.
 FT DOMAIN 465 513 LAMININ EGF-LIKE 4.
 FT DOMAIN 514 523 LAMININ EGF-LIKE 5 (N-TERMINAL).
 FT DOMAIN 524 719 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 720 752 LAMININ EGF-LIKE 5 (C-TERMINAL).
 FT DOMAIN 753 802 LAMININ EGF-LIKE 6.
 FT DOMAIN 803 860 LAMININ EGF-LIKE 7.
 FT DOMAIN 861 913 LAMININ EGF-LIKE 8.
 FT DOMAIN 914 962 LAMININ EGF-LIKE 9.
 FT DOMAIN 963 1009 LAMININ EGF-LIKE 10.
 FT DOMAIN 1010 1055 LAMININ EGF-LIKE 11.
 FT DOMAIN 1056 1101 LAMININ EGF-LIKE 12.
 FT DOMAIN 1102 1161 LAMININ EGF-LIKE 13.
 FT DOMAIN 1162 1171 LAMININ EGF-LIKE 14 (N-TERMINAL).
 FT DOMAIN 1172 1375 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 FT DOMAIN 1376 1415 LAMININ EGF-LIKE 14 (C-TERMINAL).
 FT DOMAIN 1416 1464 LAMININ EGF-LIKE 15.
 FT DOMAIN 1465 1522 LAMININ EGF-LIKE 16.
 FT DOMAIN 1523 1569 LAMININ EGF-LIKE 17.
 FT DOMAIN 1570 2140 DOMAIN II AND I.
 FT DOMAIN 2141 2324 LAMININ G-LIKE 1.
 FT DOMAIN 2325 2376 LAMININ G-LIKE 2.
 FT DOMAIN 2377 2706 LAMININ G-LIKE 3.
 FT DOMAIN 2707 2930 LAMININ G-LIKE 4.
 FT DOMAIN 2931 3106 LAMININ G-LIKE 5.
 FT DOMAIN 3107 1863 COILED COIL (POTENTIAL).
 FT DOMAIN 1864 2146 COILED COIL (POTENTIAL).
 FT DISULFID 283 292 BY SIMILARITY.
 FT DISULFID 285 303 BY SIMILARITY.
 FT DISULFID 305 314 BY SIMILARITY.
 FT DISULFID 317 337 BY SIMILARITY.
 FT DISULFID 340 349 BY SIMILARITY.
 FT DISULFID 342 374 BY SIMILARITY.
 FT DISULFID 377 386 BY SIMILARITY.
 FT DISULFID 389 407 BY SIMILARITY.
 FT DISULFID 410 422 BY SIMILARITY.
 FT DISULFID 412 438 BY SIMILARITY.
 FT DISULFID 440 449 BY SIMILARITY.
 FT DISULFID 452 462 BY SIMILARITY.
 FT DISULFID 465 478 BY SIMILARITY.
 FT DISULFID 467 482 BY SIMILARITY.
 FT DISULFID 484 493 BY SIMILARITY.
 FT DISULFID 496 511 BY SIMILARITY.
 FT DISULFID 511 762 BY SIMILARITY.
 FT DISULFID 753 762 BY SIMILARITY.
 FT DISULFID 755 769 BY SIMILARITY.
 FT DISULFID 772 781 BY SIMILARITY.
 FT DISULFID 784 800 BY SIMILARITY.
 FT DISULFID 803 818 BY SIMILARITY.
 FT DISULFID 805 828 BY SIMILARITY.
 FT DISULFID 831 840 BY SIMILARITY.
 FT DISULFID 843 858 BY SIMILARITY.
 FT DISULFID 861 875 BY SIMILARITY.
 FT DISULFID 883 882 BY SIMILARITY.
 FT DISULFID 885 894 BY SIMILARITY.
 FT DISULFID 897 911 BY SIMILARITY.
 FT DISULFID 914 926 BY SIMILARITY.
 FT DISULFID 916 933 BY SIMILARITY.
 FT DISULFID 935 944 BY SIMILARITY.
 FT DISULFID 947 960 BY SIMILARITY.
 FT DISULFID 963 975 BY SIMILARITY.
 FT DISULFID 965 981 BY SIMILARITY.
 FT DISULFID 983 992 BY SIMILARITY.
 FT DISULFID 995 1007 BY SIMILARITY.
 FT DISULFID 1010 1019 BY SIMILARITY.
 FT DISULFID 1012 1026 BY SIMILARITY.
 FT DISULFID 1028 1037 BY SIMILARITY.

FT DISULFID 1040 1053 BY SIMILARITY.
 FT DISULFID 1056 1068 BY SIMILARITY.
 FT DISULFID 1075 1086 BY SIMILARITY.
 FT DISULFID 1077 1086 BY SIMILARITY.
 FT DISULFID 1089 1099 BY SIMILARITY.
 FT DISULFID 1416 1425 BY SIMILARITY.
 FT DISULFID 1418 1432 BY SIMILARITY.
 FT DISULFID 1435 1444 BY SIMILARITY.
 FT DISULFID 1447 1462 BY SIMILARITY.
 FT DISULFID 1465 1480 BY SIMILARITY.
 FT DISULFID 1467 1490 BY SIMILARITY.
 FT DISULFID 1493 1502 BY SIMILARITY.
 FT DISULFID 1505 1520 BY SIMILARITY.
 FT DISULFID 1523 1535 BY SIMILARITY.
 FT DISULFID 1525 1542 BY SIMILARITY.
 FT DISULFID 1544 1553 BY SIMILARITY.
 FT DISULFID 1556 1567 BY SIMILARITY.
 FT DISULFID 1570 1570 INTERCHAIN (PROBABLE).
 FT DISULFID 1574 1574 INTERCHAIN (PROBABLE).
 FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 6.0%; Score 98.5; DB 1; Length 3106;
 Best Local Similarity 22.0%; Pred. No. 2.4;
 Matches 70; Conservative 21; Mismatches 82; Indels 145; Gaps 19;

Qy 73 PSVNGSAPP-----SGSTL---PGPPKITLAGYDCDFAS---GDPC 108
 Db 851 PVPGGSCQPCQCNLDYSPGSCDSUSGCLCKPG---TTGRYELCALGYFGDAV 906
 Qy 109 NNCNCCN-CN-----NLHHDIERFAIKRAICLGRN-----PEAFQPKIGKQGLGNVKP 155
 Db 907 NTKNCQPCRCDSINGSFSDCHTRTQCCECPNVQGRHCDCKPETFGLQGRGCL----- 961
 Qy 156 QHNKGCNRRSGCLKNYCEYEAQIMCS---SICKICGCK-----NYEE----- 196
 Db 962 ---PCNCSNFGSKSFDCGA-SGQCWCQPGVAGKCKDCRCAHGYNFQBGCGCIACDCSHLG 1016
 Qy 197 ---SPERKTLMSMPNYMOTGLEGSHYLPPT----- 224
 Db 1017 NNCDPKTQCICPPN---TTGEKCECPNWTGHSIVTGCKVCNCSTVGSLASQCNCNVNTG 1073
 Qy 225 ---KFSGLPRFSDRRPSSCI---SWEV-----VEAT-----CACLLA 256
 Db 1074 QCSCHPKFSGM-----KCECSRGHNYPLCTLCDCFLPGTDATTCDTLETRKCSQSDQ 1126
 Qy 257 QGE-----EAEKEHCKSC 269
 Db 1127 TQCCKVNVGVHCDCR 1144

RESULT 15
 ZAN_MOUSE
 ID ZAN_MOUSE STANDARD; PRT; 5376 AA.
 AC O88799; O08647;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zonadhesin precursor.
 GN ZAN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=98123114; PubMed=9452463;

RA Gao Z., Garbers D.L.;
 RT "Species diversity in the structure of zonadhesin, a sperm-specific
 RT membrane protein containing multiple cell adhesion molecule-like
 RL domains";
 RL J. Biol. Chem. 273:3415-3421(1998).
 RN [2]
 RP SEQUENCE OF 4864-5376 FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=97271566; PubMed=9126492;
 RA Gao Z., Harumi T., Garbers D.L.;
 RT "Chromosome localization of the mouse zonadhesin gene and the human
 RT zonadhesin gene (ZAN).";
 RL Genomics 41:119-122(1997).
 CC -1- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
 CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
 CC SIGNALING.
 CC -1- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
 CC APICAL REGION OF THE SPERM HEAD.
 CC -1- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
 CC -1- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
 CC ZONA PELLUCIDA.
 CC -1- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
 CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
 CC SPERMATOZOEA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
 CC -1- DOMAIN: THE VWFD DOMAIN 2 MAY MEDIATE COVALENT
 CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
 CC -1- SIMILARITY: CONTAINS 3 MAM DOMAINS.
 CC -1- SIMILARITY: CONTAINS 25 VWFD DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U97068; AAC5680.1; -;
 CC EMBL; U83190; AAC53125.1; -;
 CC MGD; MGI:106656; zan.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR003645; FOLN.
 CC InterPro; IPR000998; MAM.
 CC InterPro; IPR002919; TIL.
 CC InterPro; IPR003328; TILA.
 CC InterPro; IPR001007; VWFC.
 CC InterPro; IPR001846; Vwd.
 CC Pfam; PF00629; MAM; 3.
 CC Pfam; PF01826; TIL; 25.
 CC Pfam; PF02345; TILA; 25.
 CC Pfam; PF00094; vwd; 4.
 CC SMART; SM00181; EGF; 2.
 CC SMART; SM00274; FOLN; 11.
 CC SMART; SM00137; MAM; 2.
 CC SMART; SM00214; VWC; 17.
 CC SMART; SM00216; VWD; 4.
 CC PROSITE; PS00022; EGF_1; 1.
 CC PROSITE; PS01186; EGF_2; 18.
 CC PROSITE; PS00740; MAM_1; FALSE_NEG.
 CC PROSITE; PS00060; MAM_2; 3.
 CC Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
 KW Repeat.
 FT SIGNAL
 FT CHAIN 1 17 POTENTIAL.
 FT ZONADHESIN.
 FT DOMAIN 18 5376 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 18 5310 POTENTIAL.
 FT DOMAIN 5311 5337 CYTOPLASMIC (POTENTIAL).
 FT MAM 1.
 FT DOMAIN 45 210 MAM 1.
 FT DOMAIN 215 374 MAM 2.
 FT DOMAIN 377 542 MAM 3.
 FT DOMAIN 547 1170 80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)

FT DOMAIN 1171 1280 (MUCIN-LIKE DOMAIN).
 FT VWFD 1 (PARTIAL).
 FT DOMAIN 1281 1669
 FT VWFD 2.
 FT DOMAIN 1670 2056
 FT VWFD 3.
 FT DOMAIN 2057 2459
 FT VWFD 4.
 FT DOMAIN 2460 2579
 FT VWFD 5 (PARTIAL).
 FT DOMAIN 2580 2699
 FT VWFD 6 (PARTIAL).
 FT DOMAIN 2700 2819
 FT VWFD 7 (PARTIAL).
 FT DOMAIN 2820 2939
 FT VWFD 8 (PARTIAL).
 FT DOMAIN 2940 3059
 FT VWFD 9 (PARTIAL).
 FT DOMAIN 3060 3179
 FT VWFD 10 (PARTIAL).
 FT DOMAIN 3180 3299
 FT VWFD 11 (PARTIAL).
 FT DOMAIN 3300 3416
 FT VWFD 12 (PARTIAL).
 FT DOMAIN 3417 3536
 FT VWFD 13 (PARTIAL).
 FT DOMAIN 3537 3656
 FT VWFD 14 (PARTIAL).
 FT DOMAIN 3657 3776
 FT VWFD 15 (PARTIAL).
 FT DOMAIN 3777 3892
 FT VWFD 16 (PARTIAL).
 FT DOMAIN 3893 4928
 FT VWFD 17 (PARTIAL).
 FT DOMAIN 4029 4148
 FT VWFD 18 (PARTIAL).
 FT DOMAIN 4149 4263
 FT VWFD 19 (PARTIAL).
 FT DOMAIN 4264 4283
 FT VWFD 20 (PARTIAL).
 FT DOMAIN 4384 4503
 FT VWFD 21 (PARTIAL).
 FT DOMAIN 4504 4623
 FT VWFD 22 (PARTIAL).
 FT DOMAIN 4624 4743
 FT VWFD 23 (PARTIAL).
 FT DOMAIN 4744 4863
 FT VWFD 24 (PARTIAL).
 FT DOMAIN 4864 5261
 FT VWFD 25.
 FT EGF-LIKE.
 FT BY SIMILARITY.
 FT DISULFID 5263 5274
 FT BY SIMILARITY.
 FT DISULFID 5268 5283
 FT BY SIMILARITY.
 FT DISULFID 5285 5294
 FT BY SIMILARITY.
 FT CARBOHYD 339 339
 FT CARBOHYD 499 499
 FT CARBOHYD 1216 1216
 FT CARBOHYD 1239 1239
 FT CARBOHYD 1314 1314
 FT CARBOHYD 1814 1814
 FT CARBOHYD 1908 1908
 FT CARBOHYD 1933 1933
 FT CARBOHYD 2028 2028
 FT CARBOHYD 2111 2111
 FT CARBOHYD 2142 2142
 FT CARBOHYD 2332 2332
 FT CARBOHYD 2533 2533
 FT CARBOHYD 2575 2575
 FT CARBOHYD 2692 2692
 FT CARBOHYD 2812 2812
 FT CARBOHYD 3052 3052
 FT CARBOHYD 3065 3065
 FT CARBOHYD 3144 3144
 FT CARBOHYD 3172 3172
 FT CARBOHYD 3288 3288
 FT CARBOHYD 3292 3292
 FT CARBOHYD 3782 3782
 FT CARBOHYD 4005 4005
 FT CARBOHYD 4136 4136
 FT CARBOHYD 4243 4243
 FT CARBOHYD 4254 4254
 FT CARBOHYD 4335 4335
 FT CARBOHYD 4376 4376
 FT CARBOHYD 4586 4586
 FT CARBOHYD 5136 5136
 FT CARBOHYD 5252 5252
 FT SEQUENCE 5376 AA; 579908 MW; 0B44DB77DF2A2620 CRC64;

Query Match 6.0%; Score 98.5; DB 1; Length 5376;

Best Local Similarity 23.2%; Pred. No. 4.4;
 Matches 53; Conservative 19; Mismatches 73; Indels 83; Gaps 15;

QY 99 CDCFASGDFCNCCNC-----NNCCNNLHDIERFKAIRKACL-GRNPEAFQPKIGKQL 150

DB 4232 CACVGGVQCHNFTCTGTGTCQNSCSKITVQCPAHSQYTTCLPCLSPCDFP---EGLC 4288

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:22:18 ; Search time 92.59 seconds
(without alignments)
558.651 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647

Sequence: 1 MVICQLKGGTQMLCIDSRT.....GRCLSQLHTEFKSLKME 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL19:

- 1: sp_archaea:
- 2: sp_bacteria:
- 3: sp_fungi:
- 4: sp_human:
- 5: sp_invertebrate:
- 6: sp_mammal:
- 7: sp_mhc:
- 8: sp_organelle:
- 9: sp_phase:
- 10: sp_plant:
- 11: sp_rodent:
- 12: sp_virus:
- 13: sp_vertebrate:
- 14: sp_unclassified:
- 15: sp_rvirus:
- 16: sp_bacteriap:
- 17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	470	28.5	950	5 Q9V6Q8	Q9V6Q8 drosophila
2	437	26.5	429	5 Q62295	Q62295 caenorhabdi
3	437	26.5	435	5 Q95QD7	Q95QD7 caenorhabdi
4	376	22.8	571	10 Q9SL70	Q9SL70 arabidopsis
5	369	22.4	277	11 Q9D571	Q9D571 mus musculus
6	357	21.7	603	10 Q9SD1	Q9SD1 arabidopsis
7	262.5	15.9	356	10 Q9LW71	Q9LW71 arabidopsis
8	261	15.8	601	10 Q9CAV1	Q9CAV1 arabidopsis
9	248.5	15.1	896	10 Q9ZS22	Q9ZS22 glycine max
10	247.5	15.0	609	10 Q9M679	Q9M679 arabidopsis
11	245.5	14.9	658	10 Q23333	Q23333 arabidopsis
12	229.5	13.9	526	10 Q94A12	Q94A12 arabidopsis
13	229.5	13.9	695	10 Q9LE32	Q9LE32 arabidopsis
14	229.5	13.9	695	10 Q9LU13	Q9LU13 arabidopsis
15	202.5	12.3	593	10 Q9LU15	Q9LU15 arabidopsis
16	200	12.1	243	5 Q9VMQ3	Q9VMQ3 drosophila

17	187.5	11.4	553	10 Q94DS2	Q94DS2 oryza sativ
18	117	7.1	497	5 Q23460	Q23460 caenorhabdi
19	113	6.9	1981	5 Q9VVK7	Q9VVK7 drosophila
20	112	6.8	1737	4 Q75097	Q75097 homo sapien
21	107.5	6.5	1282	5 Q18720	Q18720 entamoeba d
22	106	6.4	3010	12 Q9DTE7	Q9DTE7 hepatitis c
23	105.5	6.4	911	11 Q9CRX6	Q9CRX6 mus musculu
24	105	6.4	984	5 Q9Y1P7	Q9Y1P7 cryptospori
25	104	6.3	773	5 Q17514	Q17514 caenorhabdi
26	104	6.3	773	5 Q62335	Q62335 caenorhabdi
27	103.5	6.3	398	11 Q9LYN8	Q9LYN8 mus musculu
28	103.5	6.3	1125	5 P92135	P92135 entamoeba d
29	103.5	6.3	3010	12 Q9QIX2	Q9QIX2 hepatitis c
30	103	6.3	552	15 Q83934	Q83934 ovine lenti
31	102.5	6.2	4123	4 Q75851	Q75851 homo sapien
32	102	6.2	420	5 P91776	P91776 pacifastacu
33	102	6.2	552	15 Q83935	Q83935 ovine lenti
34	102	6.2	2192	5 Q01768	Q01768 caenorhabdi
35	102	6.2	2327	13 Q9IBG7	Q9IBG7 xenopus lae
36	101	6.1	552	15 Q83932	Q83932 ovine lenti
37	101	6.1	552	15 Q83933	Q83933 ovine lenti
38	101	6.1	1819	16 Q9ZLV0	Q9ZLV0 helicobacte
39	101	6.1	2436	12 Q81756	Q81756 hepatitis c
40	101	6.1	3010	12 Q9QIX8	Q9QIX8 hepatitis c
41	101	6.1	3010	12 Q9QIX7	Q9QIX7 hepatitis c
42	101	6.1	3011	12 Q9IFE5	Q9IFE5 hepatitis c
43	100.5	6.1	633	5 Q22468	Q22468 caenorhabdi
44	100.5	6.1	760	5 Q9VTA3	Q9VTA3 drosophila
45	100.5	6.1	1028	11 Q9JLL0	Q9JLL0 mus musculu

ALIGNMENTS

RESULT 1

Q9V6Q8	PRELIMINARY;	PRT;	950 AA.
ID	Q9V6Q8		
AC	Q9V6Q8:		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)		
DE	CG6061 PROTEIN.		
GN	CG6061.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-BERKELEY;		
RX	MEDLINE=20196006; PubMed=10731132;		
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abel J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Fieriera S., Fleischmann W.,		
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Hough J.,		
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		

Db 367 TMCVQAEEALNYEKVQTEDEKLNMEKLVLEFGRCLEQMI 409

RESULT 4
Q9SL70 ID Q9SL70 PRELIMINARY; PRT; 571 AA.
AC Q9SL70;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE AT2G20110 PROTEIN.
GN AT2G20110.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana";
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006081; AAD24386.1;
SQ SEQUENCE 571 AA; 61312 MW; 36701C5FA62B3B89 CRC64;

Query Match 22.8%; Score 376; DB 10; Length 571;
Best Local Similarity 29.4%; Pred. No. 4.1e-31;
Matches 95; Conservative 48; Mismatches 90; Indels 90; Gaps 10;
QY 41 SDVPKPMALVGRFLPASTKLNLT-----QLEGAL-----PSVYNGSPSGST 86
DB 51 ASIPSPIVTVTRPIITSQAPPTVATPIPPPSQGIILHVPIRHPRESPNSMPRAGET 110
QY 87 LFGPPK-----ITLAGYCDCAFSGDFCNCCNCCNHLHDIERFKAICACGLR 136
DB 111 RGTTPQKKQCKNCKSRCLKLYCECFASGYCDGNCVNCNVEPARRQAVESTLER 170
QY 137 NPEAFOPKI-----GKQGLGVN--KQHNKGNCRRSGLKNKYCYEQAQIMCSSI 185
DB 171 NPNAEPKIAASPHGGRDNREEVGDVVMLARHNKCKKSGCLKCYCECFQANILCSEN 230
QY 186 KCICGCKNVEESPRTKLTASMPN-----YMQ-----TGLEGSHYL---PPTKFSG--- 228
DB 231 CKCLDCKNFESEVRSOLFGEHSHNLAYLQHANAAITGAIGSSGFASAPPPKRRKQGEI 290
QY 229 -----LPRFSDHRRPSS-----CISWEVVEAT 250
DB 291 FFGNQTGKDSSTHRLGQANNRTTSQTGSRAGNAGSLGFSKVYKSLLANIKPMVDKAL 350
QY 251 CACLLAQGEAEKEHCKSLAEQ 273
DB 351 CSVLVAVAGEAAKLTLEKRLANQ 373
RESULT 5
Q9D571 ID Q9D571 PRELIMINARY; PRT; 277 AA.
AC Q9D571;

DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ADULT.MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE CLONE:4930509C02, FULL INSERT SEQUENCE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito K.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection";
RL Nature 409:685-690(2001).
DR EMBL; AK015732; BAB29949.1;
SQ SEQUENCE 277 AA; 28263 MW; 1731F517A3CE4D43 CRC64;

Query Match 22.4%; Score 369; DB 11; Length 277;
Best Local Similarity 73.2%; Pred. No. 9e-31;
Matches 71; Conservative 8; Mismatches 18; Indels 0; Gaps 0;
QY 1 MVTCQLKGGTQMLCIDNSTRRELKALHLVPOYQDQNNYQSDVPKPMALVGRFLPASTK 60
DB 181 MVTCQLKGGTQMLCIDNSTRRELKALHLVPOYDQDQNNYQSDVPKPMALVGRFLPASTK 240
QY 61 LNLITQQLSCALPSVYNGSPSGSTLPGPKITLAG 97
DB 241 LNLITQDNGALPSAYNGAAPPSPGALQGPPTLTS 277
RESULT 6
Q9SZD1 ID Q9SZD1 PRELIMINARY; PRT; 603 AA.
AC Q9SZD1;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR.
GN F19B15.30 OR AT4G29000.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

Best Local Similarity 29.7%; Pred. No. 6.7e-19;
Matches 69; Conservative 25; Mismatches 64; Indels 74; Gaps 10;

QY 33 ODNNYLOSD-----VPKPWAL-----VGRFLPASTKLNLIQOLEGAL-PSV----- 75
DB 307 RDSNDLPDSTSIKAPSPONCLDTSKQDTDEILPRTIGL-----HLNGFVNPVSNGR 363

QY 76 -----VNGSAPFSGS-----TLPGPKITLAG----- 97
DB 364 KKKIKQQAFTTFHYNIEDEFSTVSTKRDLVFSDVKIMEPPERSVEGECDQLMA 423

QY 98 ----YDCDFASGDFCNN-CNCNCCNNLHHDIERFKAICACLRNPEAFQPKIGKG--- 149
DB 424 MENRYCEFSAGLFCGEGPCSCQCNCFKPIHEDLVNKSREVIKARNPLAFAPKVVSTSDTV 483

QY 150 ----LGNVK-----POHNGKCNRRSGCLNKYCEYEAQIMCSSICKIGCKN 193
DB 484 IDLWVNSKTPASARHKRGCKNRKSGCKKYCECFMGMVGCCSNCRMGCKN 535

RESULT 9
Q92S22 ID Q92S22 PRELIMINARY; PRT; 896 AA.
AC Q92S22

DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE CYSTEINE-RICH POLYCOMB-LIKE PROTEIN.
GN CPPL.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=20345129; PubMed=10859345;
RA Cvitanič C., Pallisgaard N., Nielsen K.A., Hansen A.C., Larsen K.,
RA Pihakeski-Maunsbach K., Marcker K.A., Jensen E.O.;
RT "CPL, a novel type DNA-binding protein involved in the expression of
a soybean leghemoglobin c3 gene";
RL Proc. Natl. Acad. Sci. U.S.A. 97:8163-8168(2000).
DR EMBL: AJ010165; CAA09028.1; -
SQ SEQUENCE 896 AA; 97488 MW; 6E27E1B5E4D98DC7 CRC64;

Query Match 15.1%; Score 248.5; DB 10; Length 896;
Best Local Similarity 40.7%; Pred. No. 2.3e-17;
Matches 50; Conservative 13; Mismatches 31; Indels 29; Gaps 5;

QY 98 YDCDFASGDFCNN-CNCNCCNNLHHDIERFKAICACLRNPEAFQPKIGKGNVVKP 155
DB 487 YDCDFASGDFCNN-CNCNCCNNLHHDIERFKAICACLRNPEAFQPKIGKGNVVKP 538

QY 156 -----OHNKGCNRRSGCLNKYCEYEAQIMCSSICKIGCKNYEE 196
DB 539 TTDISSHMDENLTTPSSARHKRGCKNRKSGCKKYCECFMGMVGCCSNCRMGCKN 598

QY 197 SPE 199
DB 599 KKE 601

RESULT 10
Q9M679 ID Q9M679 PRELIMINARY; PRT; 609 AA.
AC Q9M679;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CXC DOMAIN CONTAINING TSO1-LIKE PROTEIN 1.

GN SOL1.
OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]

RP SEQUENCE FROM N.A.
RX STRAIN=CV. LER;
RA Hausner B.A., He J., Park S.O., Gasser C.S.;

RT "TSO1 is a novel protein that modulates cytokinesis and cell expansion
in Arabidopsis";
RL Development 127:2219-2226(2000).
DR EMBL: AF205142; AAF69125.1; -
SQ SEQUENCE 609 AA; 66738 MW; 7E29C055A2423089 CRC64;

Query Match 15.0%; Score 247.5; DB 10; Length 609;
Best Local Similarity 38.6%; Pred. No. 1.8e-17;
Matches 49; Conservative 20; Mismatches 35; Indels 23; Gaps 6;

QY 98 YDCDFASGDFCNN-CNCNCCNNLHHDIERFKAICACLRNPEAFQPKIGK----- 147
DB 341 YCEFAAGFYCIETPCSCNCFNPKPIHKDVV-LATRKQIESRNPAPAFAPKVRNSDSIIEV 399

QY 148 GOLGNVVKP---QHNKGCNRRSGCLNKYCEYEAQIMCSSICKIGCKN-----YE 195
DB 400 GEDASKTPASARHKRGCKNRKSGCKKYCECFMGMVGCCSNCRMGCKN 459

QY 196 ESPERKT 202
DB 460 QDEENET 466

RESULT 11
O23333 ID O23333 PRELIMINARY; PRT; 658 AA.
AC O23333;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 72.1 KDA PROTEIN.
GN AT4G14770.

OC Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=98121113; PubMed=9461215;

RA Bevan M., Bancroft I., Bent E., Love K., Goodman H., Dean C.,
RA Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L.,
RA Ridley P., Hudson S.A., Patel K., Murphy G., Piffanelli P., Wedler H.,
RA Wedler E., Wambutt R., Weitzenegger T., Pohl T.M., Terry N.,
RA Gielen J., Villarroel R., De Clerck R., van Montagu M., Lecharny A.,
RA Auborg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., Kottler P.,
RA Entian K.D., Rieger M., Schaeffer M., Funk B., Mueller-Auer S.,
RA Silvey M., James R., Montfort A., Pons A., Puigdomenech P., Douka A.,
RA Vouklatou E., Milioni D., Hatzopoulos P., Piravandi E., Obermaier B.,
RA Hilbert H., Duesterhoft A., Moeres T., Jones J.D.G., Eneva T.,
RA Palme K., Benes V., Rechman S., Ansorge W., Cooke R., Berger C.,
RA Delsen M., Voet M., Volckaert G., Mewes H.W., Klosterman S.,
RA Schueller C., Chaltatzis N.

RT "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
Arabidopsis thaliana";
RL Nature 391:485-488(1998).
RN [2]

RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z97337; CAB10256.1; -
DR EMBL: AL161539; CAB78519.1; -
KW Hypothetical protein.

Search completed: July 11, 2002, 08:22:19
Job time: 478 sec

Query Match 13.9%; Score 229.5; DB 10; Length 695;
Best Local Similarity 24.2%; Pred. No. 1.7e-15;
Matches 55; Conservative 31; Mismatches 72; Indels 69;

RESULT 15
Q9LUI5
ID Q9LUI5 PRELIMINARY; PRT; 593 AA.

Query Match 12.3%; Score 202.5; DB 10; Length 593;
Best Local Similarity 36.1%; Pred. No. 1e-12;
Matches 43; Conservative 21; Mismatches 32; Indels 23;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:16:00 ; Search time 34.45 Seconds
(without alignments)
211.996 Million cell updates/sec

Title: US-09-743-237-5

Perfect score: 1647

Sequence: 1 MVICQLKGGTQMLCIDNSRT.....GRCLSOILHTEFKSKGLKWE 299

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued_Patents_AA:*
- 1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
 - 2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
 - 3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
 - 4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
 - 5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
 - 6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	109.5	6.6	769	2	US-08-789-078-1
2	109.5	6.6	769	2	US-08-752-633-1
3	109.5	6.6	769	2	US-08-476-062A-45
4	109.5	6.6	769	2	US-07-728-215-31
5	109.5	6.6	769	4	US-08-938-085A-31
6	109.5	6.6	769	5	PCT-US95-04886-1
7	109.5	6.6	769	5	PCT-US96-01314-45
8	108	6.6	676	3	US-08-630-172-10
9	108	6.6	676	4	US-09-375-419-10
10	108	6.6	739	4	US-08-444-818-148
11	108	6.6	2995	4	US-08-444-818-138
12	101	6.1	465	2	US-08-833-678A-2
13	101	6.1	465	4	US-08-529-169A-2
14	101	6.1	631	2	US-08-833-678A-1
15	101	6.1	631	4	US-08-529-169A-1
16	101	6.1	1021	1	US-07-910-760-12
17	101	6.1	1021	1	US-08-440-519-12
18	101	6.1	1021	4	US-08-440-549-12
19	101	6.1	1786	4	US-08-444-818-54
20	101	6.1	2261	4	US-08-444-818-66
21	101	6.1	2436	4	US-08-444-818-75
22	101	6.1	2772	4	US-08-444-818-89
23	101	6.1	2894	2	US-08-466-975A-23
24	101	6.1	2894	2	US-08-391-671A-23
25	101	6.1	2894	3	US-08-467-902A-23
26	101	6.1	2894	4	US-09-275-265-23
27	101	6.1	3011	1	US-08-440-103-36

28	101	6.1	3011	1	US-08-440-542-36	Sequence 36, Appl
29	101	6.1	3011	1	US-07-910-760-10	Sequence 10, Appl
30	101	6.1	3011	1	US-08-440-519-10	Sequence 10, Appl
31	101	6.1	3011	1	US-08-231-368-36	Sequence 36, Appl
32	101	6.1	3011	1	US-08-440-210-36	Sequence 36, Appl
33	101	6.1	3011	4	US-09-388-874-2	Sequence 2, Appl
34	101	6.1	3011	4	US-09-046-604-36	Sequence 36, Appl
35	101	6.1	3011	4	US-08-440-549-10	Sequence 10, Appl
36	97.5	5.9	969	2	US-08-284-941-2	Sequence 2, Appl
37	97.5	5.9	969	2	US-08-447-642-2	Sequence 2, Appl
38	97.5	5.9	969	4	US-09-236-503-2	Sequence 2, Appl
39	97.5	5.9	969	5	PCT-US93-02147A-2	Sequence 2, Appl
40	97	5.9	2955	2	US-08-443-260-3	Sequence 3, Appl
41	97	5.9	2955	3	US-08-442-805A-3	Sequence 3, Appl
42	97	5.9	2955	3	US-08-443-900A-3	Sequence 3, Appl
43	97	5.9	2955	4	US-08-444-818-124	Sequence 124, App
44	97	5.9	2955	4	US-08-249-843-3	Sequence 3, Appl
45	97	5.9	3011	2	US-08-833-678A-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-08-789-078-1
; Sequence 1, Application US/08789078
; Patent No. 5843885
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.
; APPLICANT: Siahaan, Teruna
; APPLICANT: Chan, Marcia
; APPLICANT: Tibbetts, Scott
; TITLE OF INVENTION: ICAM-1/LEA-1 PEPTIDES FOR INDUCING
; TITLE OF INVENTION: IMMUNE TOLERANCE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 1101 Walnut St.
; CITY: Kansas City
; STATE: MO
; COUNTRY: USA
; ZIP: 64106
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/789,078
; FILING DATE: 03-FEB-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/229,513
; FILING DATE: 19-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26262
; REFERENCE/DOCKET NUMBER: 22833
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (816)474-9050
; TELEFAX: 816)474-9057
; TELEX: 434-363
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 769 amino acids
; TYPE: amino acid
; TOPOLOGY: both
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Tonsil
; FEATURE:

NAME/KEY: Region
LOCATION: 1..22 /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496 /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 497..540 /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581 /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617 /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723 /label= trans
OTHER INFORMATION: /note= "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769 /label= cyto
OTHER INFORMATION: /note= "cytoplasmic domain"
PUBLICATION INFORMATION:
AUTHORS: Pigott,
TITLE: LFA-1 Amino acid sequence (B2) (from human
JOURNAL: The Adhesion Molecule Facts Book
PAGES: 96-96
DATE: 1993
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 769
US-08-789-078-1

Query Match 6.6%; Score 109.5; DB 2; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.029;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;
QY 1 MVICQLKGGTQMLCIDNSRTRELKALHLVPQYQDNNYLOSDVPKPM TALVGRFLPASTK 60
Db 436 IVTVVLPQCECRDQSRSL--CH-----GKFLGEGICRCDTGVIGNCECOTQ 486
QY 61 LNLITQOLEG-----LPSVYNGSAFPGSTLPFPKPTILAGYCDGFA----- 103
Db 487 -GRSSQELGEGSCRDNNSIICGLGDCVCGCLHTSDVPG--KLIYQYCECDTINCR 543
QY 104 -SGDFCNN-----CNCNCCNLLHHDIERF-----KAIKACLGNPFAFPKIGKGLGN 152
Db 544 YNGQVCGGPGRLGFCGKCK--RCHPGFEGSACOCERTTEGL--NPRRVECS-GRGR--- 595
QY 153 VKPOHNGKCNRRSGGLKNTCECYEAQIM-----CSSITC-----KCIGCKNYEESP 198
Db 596 -----CRC-----NVCECHSGQLPLCOECPCGSPCGKIYISCAELKFEKGP 638

RESULT 2
US-08-752-633-1
; Sequence 1, Application US/08752633
; Patent No. 586389
; GENERAL INFORMATION:
; APPLICANT: Benedict, Stephen H.

APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
TITLE OF INVENTION: IMMUNE TOLERANCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,633
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELECOMMUNICATION INFORMATION:
TELEPHONE: (816)474-9050
TELEFAX: 816)474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Tonsil
FEATURE:
NAME/KEY: Region
LOCATION: 1..22 /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496 /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 497..540 /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 541..581 /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Region
LOCATION: 582..617 /label= repeat
OTHER INFORMATION: /note= "cysteine rich repeat"
FEATURE:
NAME/KEY: Domain
LOCATION: 701..723 /label= trans
OTHER INFORMATION: /note= "transmembrane domain"
FEATURE:
NAME/KEY: Domain
LOCATION: 724..769

TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-476-062A-45

Query Match 6.6%; Score 109.5; DB 2; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.029;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY	1	MYLCJQUGGTQMLC	INDNSTRELKALHLVPODQNNY	LDSDVPRKPMITALVGRFLPASTK	60	
		: : : : : :	: : : : : :	: : : : : :		
Db	436	IVTVQLPQCECRCDQSDRSL	-CH-----	KGFLCEGICRCDTGYICKNCECQTQ	486	
		: : : : : :				
QY	61	LNLIQTQLGGA	-----	LPSVWNGSAPPSGSTLPGPPKTILAGYCDCA	103	
		: : : : : :				
Db	487	-GRSSQLEEGSCRKDNNSIIC	SGLGCVCVQCGLCHTSDVPG-	KLIYGYCYCEDTINCR	543	
		: : : : : :				
QY	104	-SGDFCNN	-----	CNCNCCNLLHHDIERP	152	
		: : : : : :				
Db	544	YNGQVGCGPGRGLFCGKC-	RCHPGFEGSACOCERTTEGCL-	NPRRVECS-GRGR	595	
		: : : : : :				
QY	153	VRPQHMKGCNRRSGCLKNY	CYCYEAQIM-----	CSSIC-----	KLIGCKNYDESP	198
		: : : : : :				
Db	596	-----CRC	-----	NVCCHSGYQLPQCEQCGPC	SPGCKYTISCAECLKFKPG	638
		: : : : : :				

RESULT 4
US-07-728-215-31
: Sequence 31, Application US/07728215
: Patent No. 5962643
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Dean
: APPLICANT: Quaranta, Vito
: APPLICANT: Pytela, Robert
: TITLE OF INVENTION: A No. 5962643el Integrin Beta Subunit and Uses
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: United States of America
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07728,215
: FILING DATE: 19910711
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P31 8717
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 31:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 769 amino acids
: TYPE: AMINO ACID
: STRANDEDNESS: single

TOPOLOGY: linear
US-07-728-215-31

Query Match 6.6%; Score 109.5; DB 2; Length 769;
Best Local Similarity 22.3%; Pred. No. 0.029;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVTICQLKGGTOMLCIDNSRRELKALHLVPOYQDQNNYLQSDVPKPM TALVGRFLPASTK 60
Db 436 IVTVQVLPOQCECRDQSRDSL--CH-----GKFLGCGICRCDTGYIGNKCECQTQ 486
QY 61 LNLITQOLEGA-----LPSVWVGSAPFSGSTLPGPKITLAGYCDGCF-----103
Db 487 -GRSSQLEGGSCRRKDNNSIICSLGDCVCGQCLCHTSDVPG--KLIYQYCECDTINCER 543
QY 104 -SGDFCANN-----CNCNCCNNLHHDIERF-----KAIKACLGRLNPEAFQPKIGKGLGN 152
Db 544 YNGQVCGGPRGLCFGCKC--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR---595
QY 153 VKPOHNKGCNRRSGCLKNYCEYEAQIM-----CSSIC-----KCIGCKNVEESP 198
Db 596 -----CRC-----NVCECHSGYQLPQCPCPSPCGKYISCAECLAFKGP 638

RESULT 5
US-08-938-085A-31
Sequence 31, Application US/08938085A
Patent No. 6339148
GENERAL INFORMATION:
APPLICANT: Sheppard, Dean
APPLICANT: Quaranta, Vito
APPLICANT: Pytela, Robert
TITLE OF INVENTION: A No. 6339148el Integrin Beta Subunit and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,085A
FILING DATE: 26-SEP-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/728,215
FILING DATE: 11-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Parent, Annette S.
REGISTRATION NUMBER: 42,058
REFERENCE/DOCKET NUMBER: 023070-080210US
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-938-085A-31

Query Match 6.6%; Score 109.5; DB 4; Length 769;

Best Local Similarity 22.3%; Pred. No. 0.029;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVTICQLKGGTOMLCIDNSRRELKALHLVPOYQDQNNYLQSDVPKPM TALVGRFLPASTK 60
Db 436 IVTVQVLPOQCECRDQSRDSL--CH-----GKFLGCGICRCDTGYIGNKCECQTQ 486
QY 61 LNLITQOLEGA-----LPSVWVGSAPFSGSTLPGPKITLAGYCDGCF-----103
Db 487 -GRSSQLEGGSCRRKDNNSIICSLGDCVCGQCLCHTSDVPG--KLIYQYCECDTINCER 543
QY 104 -SGDFCANN-----CNCNCCNNLHHDIERF-----KAIKACLGRLNPEAFQPKIGKGLGN 152
Db 544 YNGQVCGGPRGLCFGCKC--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR---595
QY 153 VKPOHNKGCNRRSGCLKNYCEYEAQIM-----CSSIC-----KCIGCKNVEESP 198
Db 596 -----CRC-----NVCECHSGYQLPQCPCPSPCGKYISCAECLAFKGP 638

RESULT 6
PCT-US95-04886-1
Sequence 1, Application PC/TUS9504886
GENERAL INFORMATION:
APPLICANT: Benedict, Stephen H.
APPLICANT: Siahaan, Teruna
APPLICANT: Chan, Marcia
APPLICANT: Tibbetts, Scott
TITLE OF INVENTION: ICAM-1/LFA-1 PEPTIDES FOR INDUCING
IMMUNE TOLERANCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
STREET: 1101 Walnut St.
CITY: Kansas City
STATE: MO
COUNTRY: USA
ZIP: 64106
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Collins, John M.
REGISTRATION NUMBER: 26262
REFERENCE/DOCKET NUMBER: 22833
TELEPHONE: (816) 474-9050
TELEFAX: 816) 474-9057
TELEX: 434-363
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 769 amino acids
TYPE: amino acid
TOPOLOGY: both
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Tonsil
FEATURE:
NAME/KEY: Region
LOCATION: 1..22
OTHER INFORMATION: /label= signal
OTHER INFORMATION: /note= "signal sequence"
FEATURE:
NAME/KEY: Region
LOCATION: 449..496

```
Query Match          6.6%; Score 109.5; DB 5; Length 769;  
Best Local Similarity 22.3%; Pred. No. 0.029;  
Matches      53; Conservative    26; Mismatches   84; Indels    75; Gaps     15;
```

QY

1 MWICQLKGGTQMCLIDNSRTRRLKALHLVPOYDQNYYLQSDDVPKPMTALVGRELPASTK 60

436 IIVGVGVIPOCEFCBPDSNPSI--CH-----CKHEFLCCCTCPCPTGYTKWCPQTQ 496
DB : : : : : | : :

```

RESULT      8.
US-08-630-172-10
; Sequence 10, Application US/08630172
; Patent No. 5050054
; GENERAL INFORMATION:
; APPLICANT: Staerz, Uwe
; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
; TITLE OF INVENTION: LYMPHOCYTE VETO
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, 35th Floor
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/630,172
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2879-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 676 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PS-08-630-172-10

```

Query Match	6.68;	Score 108;	DB 3;	Length 676;
Best local Similarity	21.5%;	Pred. No. 0.034;		
Matches	51;	Conservative	26;	Mismatches 86; Indels 74; Gaps 14;
Qy	1	MVICQLKGGTQMLCIDNSRTRELKALHLVPQYQDONNYLQSDVPKPMTALVGRFLPASTK	60	
Db	414	IVTVQVLPOCECRCDQSRSL--CH-----GKGFLEGGICRCDQYIGKNCECOTQ	464	
Qy	61	LNLIITQOLEGA-----LPSVVGVSAPFSGSTLPGPPKITLAGYCDCA-----	103	
Db	465	-GRSSQELEGSGCKRDNNSIICGLGDCVCGQCLCHTSDVPG--KLIIQYCYECDDINGER	521	
Qy	104	-SGDFCNN-----CNCNCCNLLHHIDIERF-----KATKACILGRNPEAFQPKIGKGQLGN	152	
Db	522	YNGQVCGGPGRGLFCFGKC--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR---	573	
Qy	153	VKPOHNKGCNRRSGCLKNYCECYEAQIM-----CSSICKICGCKNYESSP	198	
Db	574	-----CRC-----NVCECHSGYQLPLCQECGPGSCGKYITSCAECLKFERG	615	

RESULT 9
US-09-375-419-10
: Sequence 10, Application US/09375419
: Patent No. 6264950
: GENERAL INFORMATION:
: APPLICANT: Staerz, Uwe
: TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
: TITLE OF INVENTION: LYMPHOCYTE VETO
: NUMBER OF SEQUENCES: 41
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheridan Ross & McIntosh
: STREET: 1700 Lincoln Street, 35th Floor
: CITY: Denver
: STATE: Colorado
: COUNTRY: U.S.
: ZIP: 80203
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/375,419
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: PRIOR APPLICATION NUMBER: 08/630,172
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary J.
: REGISTRATION NUMBER: 32,020
: REFERENCE/DOCKET NUMBER: 2879-36
: TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 676 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-375-419-10

Query Match 6.6%; Score 108; DB 4; Length 676;
Best Local Similarity 21.5%; Pred. No. 0.034;
Matches 51; Conservative .26; Mismatches 86; Indels 74; Gaps 14;

Qy	1	MVICQLKGGTOMLCIDNSRREIKALHLVPOYQOQNNYLOSVPKPMWALTALVGRFLPASTK	60
Db	414	IVTVQVLPOQECRCRQSDRSL--CH-----GKFLGECRCRDQTOYIGNKCECOTO	464
Qy	61	LNLIQTQOLEGA-----LPSVVVNGSAFPGSGTLPGPKIITLAGVYCDCA-----	103
Db	465	-GRSQELEGSCRRKDNNSIITCSGLGDCVCGQCLHTSDVFG--KLIYGVCECDTINCER	521
Qy	104	-SGDFCN-----CNCNNCCNNLHHDIERF-----RAIKACLGRNPEAFQPIKGQOLGN	152
Db	522	YNGQVCGGPGRLGCLFCGKC--RCHPGFEGSACQERTTEGCL--NPRVECS-GRGR----	573
Qy	153	VKPOHNGCNCRRSGCLNKVCEYEAQIM-----CSSICKICGCKNVEESP	198
b	574	-----CRC-----NVCECHSGYQLPLQCEPCGPCSGKYISCAECLKFEKP	615

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

Matches	56;	Conservative	20;	Mismatches	79;	Indels	70;	Gaps	12;
---------	-----	--------------	-----	------------	-----	--------	-----	------	-----

QY 79 SAFF-SGSTLPGPPKITTLAGYCDCFASGDFCNCNCNCCNNL-----HHDIERFKAI 130

US-08-833-678A-2

Matches 56; Conservative 21; Mismatches 77; Indels 72; Gaps 13;

[illegible]

Db	332	-----GCANWELTP-AETTVLRLAYMTNPGLPVQCDDHLEFPWGVFTGLTHIDAHFLSQTK	386
Qy	227	SG--LPRFSDRRPSSCISWEVVEATCACLLAQGEAEKEHCSCKL	270
Db	387	SGNLP-----YLVAYOATVCAACAOAPPPSMDMKKL	419

```

RESULT 14
US-08-833-678A-1
; Sequence 1, Application US/08833678A
; Patent No. 5989905
; GENERAL INFORMATION:
; APPLICANT: HOUGHTON, MICHAEL
; APPLICANT: CHOO, QUI-LIM
; APPLICANT: HAN, JANG
; APPLICANT: CHOE, JOONHO
; TITLE-OF INVENTION: HCV NS3 PROTEIN FRAGMENTS HAVING
; TITLE OF INVENTION: HELICASE ACTIVITY AND IMPROVED SOLUBILITY
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHIRON CORPORATION
; STREET: Intellectual Property - R440, P.O. Box 8097
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94662-8097
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/833,678A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/529,169
; FILING DATE: 15-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Harbin, Alisa A.
; REGISTRATION NUMBER: 33,895
; REFERENCE/DOCKET NUMBER: 0100.005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-3274
; TELEFAX: (510) 655-3542
; TELEX: n/a
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 631 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-833-678A-1

```

Query Match	6.1%	Score 101;	DB 2;	Length 631;
Best Local Similarity	24.8%;	Pred. No. 0.15;		
Matches	56;	Conservative 21;	Mismatches 77;	Indels 72; Gaps 13;
QY	79	SAPP-SGSTLPGPKITLAGYCDFASGDCFNCCNCCNNL-----HHIEREKAI	130	
		: : : : :		
Db	398	SVIPTSGDVVVATDALMTGY-----TGDFSVIDCMTCTQTVDFLDPTFTTETILP	452	
		: : : : :		
QY	131	KACLGRRPEAFQPKIGRGQLG---NVRPQHNGKNCRRSSGLKN--XCECYEAOIMCSSI	185	
		: : : : :		
Db	453	QDAVSRTQR--RGRTRGKGPIYRFVAPGE-----RPSGMFDSVLCECYDA-----	497	
		: : : : :		
QY	186	KCIICGNKYRESPERKTLMSPNPYMQTGG-----EG-----SHYLPTPKF	226	
		: : : : :		
Db	498	----GCWAYELTP-AETTVRLRAVTMPGLPVCOODHLFEWGVTTGTTHIDAHFLSQTKQ	552	
		: : : : :		
QY	227	SG--LPFRSHDRPPSSCSISWEVVEATCACLLAOGEEAEKEHCSCKL	270	

Db

553 SGENLP-----YLVA YQATVCARAQAAPPSPSWDOMWKCL 585

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	6.1%	Score	101;	DB	4;	Length	631;
Best Local Similarity	24.8%;	Pred.	No.	0.15;			
Matches	56;	Conservative	21;	Mismatches	77;	Indels	72;
Gaps	13;						
Qy	79	SAPP--SGSTLPGPKITLAGYCDCAFSGDFCNMCNCCNNL-----	HHDIERFKAI	130			
Ddb	398	SVIPTSGDVVVVATDALMTGY-----TGDFDSVIDCNTCTQTVDPSLDFTFTIETLTP	452				
Qy	131	KACLGRRNPEAFOPKIGKGQLG-----NVKPOHNKGNCRRSGCLKN--YCEYEYAOIMCSSI	185				
Ddb	453	QDAVSRTOR--RGRTGGRGPGIYRFVAPGE-----RPSGMFSSSVLCECYDA-----	497				
Qy	186	CKCIGCKNKEESPERKTLMSPNYMTGGI-----EG-----SHVLPPTKF	226				
Ddb	498	-----GCAWELTP-AETTVRLRAYMNTPLGVQCQDHLFEWEGVFTGLTHDAHFLSQTQK	552				
Qy	227	SG--LPRFSDRRPSSCISWEVVVEATCATCLLAQGEAEKHCCKL	270				
Ddb	553	SGENLP-----YLVAYQATVCARAQAPPSPSDWMKCL	585				

Search completed: July 11, 2002, 08:16:02

us-09-743-237-5.rai

Fri Jul 12 08:35:30 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 08:20:36 ; Search time 210.79 seconds
(without alignments)
499.275 Million cell updates/sec

Title: US-09-743-237-5
Perfect score: 1647
Sequence: 1 MVICQLKGGTQMLCIDSRT.....GRCLSOILHTEPKSKGLKWE 299

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 3502263 seqs, 351980561 residues
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/2/paa/US06_COMB.pap.*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pap.*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pap.*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pap.*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pap.*
6: /cgn2_6/ptodata/2/paa/US083_COMB.pap.*
7: /cgn2_6/ptodata/2/paa/US084_COMB.pap.*
8: /cgn2_6/ptodata/2/paa/US085_COMB.pap.*
9: /cgn2_6/ptodata/2/paa/US086_COMB.pap.*
10: /cgn2_6/ptodata/2/paa/US087_COMB.pap.*
11: /cgn2_6/ptodata/2/paa/US088_COMB.pap.*
12: /cgn2_6/ptodata/2/paa/US089_COMB.pap.*
13: /cgn2_6/ptodata/2/paa/US090_COMB.pap.*
14: /cgn2_6/ptodata/2/paa/US091_COMB.pap.*
15: /cgn2_6/ptodata/2/paa/US092_COMB.pap.*
16: /cgn2_6/ptodata/2/paa/US093_COMB.pap.*
17: /cgn2_6/ptodata/2/paa/US094_COMB.pap.*
18: /cgn2_6/ptodata/2/paa/US095_COMB.pap.*
19: /cgn2_6/ptodata/2/paa/US096_COMB.pap.*
20: /cgn2_6/ptodata/2/paa/US097_COMB.pap.*
21: /cgn2_6/ptodata/2/paa/US098_COMB.pap.*
22: /cgn2_6/ptodata/2/paa/US099_COMB.pap.*
23: /cgn2_6/ptodata/2/paa/US100_COMB.pap.*
24: /cgn2_6/ptodata/2/paa/US101_COMB.pap.*
25: /cgn2_6/ptodata/2/paa/US102_COMB.pap.*
26: /cgn2_6/ptodata/2/paa/US103_COMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1647	100.0	299	21	US-09-743-237-5
2	1251	76.0	295	21	US-09-743-237-4
3	1251	76.0	295	21	US-09-743-237-24
4	533.5	32.4	223	1	PCT-US01-08656-10693
5	470	28.5	403	26	US-60-161-932-1700
6	470	28.5	890	26	US-60-167-217-12948
7	470	28.5	950	20	US-09-614-150-12897

RESULT 1

US-09-743-237-5
; Sequence 5, Application US/09743237
; GENERAL INFORMATION:
; APPLICANT: SUGIHARA, TAKASHI
; APPLICANT: WADHWA, RENU
; APPLICANT: KAUL, SUNIL C.
; APPLICANT: MITSUI, YUJI
; TITLE OF INVENTION: TESTIS-SPECIFIC DIFFERENTIATION-REGULATORY FACTOR
; FILE REFERENCE: 084335/0127
; CURRENT APPLICATION NUMBER: US/09/743,237
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/JP99/03859
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 5
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-743-237-5

ALIGNMENTS

8	470	28.5	950	26	US-60-191-637-12935	Sequence 12935, A
9	446	27.1	147	1	PCT-US01-01329-1409	Sequence 1409, Ap
10	446	27.1	147	1	PCT-US01-01339-3988	Sequence 3988, Ap
11	446	27.1	147	21	US-09-764-891-3988	Sequence 3988, Ap
12	446	27.1	147	24	US-10-080-090-1409	Sequence 1409, Ap
13	437	26.5	438	14	US-09-087-136-13	Sequence 13, Appl
14	437	26.5	438	14	US-09-087-136-13	Sequence 13, Appl
15	437	26.5	438	16	US-09-220-091-13	Sequence 13, Appl
16	378	23.0	280	1	PCT-US01-08631-48317	Sequence 48317, A
17	376	22.8	524	21	US-09-708-427-33367	Sequence 33367, A
18	376	22.8	550	21	US-09-708-427-33366	Sequence 33366, A
19	376	22.8	571	21	US-09-708-427-33365	Sequence 33365, A
20	369	22.4	615	21	US-09-733-089-22002	Sequence 22002, A
21	369	22.4	615	21	US-09-733-089-22183	Sequence 22183, A
22	369	22.4	615	21	US-09-733-089-22184	Sequence 22184, A
23	369	22.4	615	22	US-09-816-660-22002	Sequence 22002, A
24	369	22.4	615	22	US-09-816-660-22183	Sequence 22183, A
25	369	22.4	615	22	US-09-816-660-22184	Sequence 22184, A
26	366.5	22.3	241	26	US-60-324-109-27382	Sequence 27382, A
27	357	21.7	603	19	US-09-573-655A-2154	Sequence 2154, Ap
28	357	21.7	603	19	US-09-573-655A-2152	Sequence 2152, Ap
29	357	21.7	603	21	US-09-708-427-27294	Sequence 27294, A
30	353	21.4	534	21	US-09-708-427-27295	Sequence 27295, A
31	351	21.3	518	21	US-09-708-427-27296	Sequence 27296, A
32	316.5	19.2	207	21	US-09-733-089-22013	Sequence 22013, A
33	316.5	19.2	207	22	US-09-816-660-22013	Sequence 22013, A
34	314.5	19.1	168	21	US-09-733-089-21980	Sequence 21980, A
35	314.5	19.1	168	22	US-09-816-660-21980	Sequence 21980, A
36	278	16.9	53	1	PCT-US00-26524B-6922	Sequence 6922, Ap
37	263	16.0	349	18	US-09-428-944-1881	Sequence 1881, Ap
38	257.5	15.6	386	26	US-60-324-109-16915	Sequence 16915, A
39	257.5	15.6	556	26	US-60-324-109-32886	Sequence 32886, A
40	245.5	14.9	497	21	US-09-708-427-25631	Sequence 25631, A
41	245.5	14.9	542	21	US-09-708-427-25630	Sequence 25630, A
42	245.5	14.9	658	21	US-09-708-427-25629	Sequence 25629, A
43	245	14.9	770	26	US-60-324-109-17053	Sequence 17053, A
44	245	14.9	770	26	US-60-324-109-17063	Sequence 17063, A
45	242.5	14.7	402	26	US-60-324-109-31233	Sequence 31233, A

Query Match 100.0%; Score 1647; DB 21; Length 299;
Best Local Similarity 100.0%; Pred. No. 6.5e-148;
Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVICQLKGGTQMLCIDSRTRELKALHLVPPQYQDQNNYQSDVPKPMALVGRFLPASTK 60
DB 1 MVICQLKGGTQMLCIDSRTRELKALHLVPPQYQDQNNYQSDVPKPMALVGRFLPASTK 60


```

QY   250 TCACLLAAGGEAEKCHSKCLAEQMILEEFRCRLSQIL 287
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db   894 TIMCMISRIYVHHEKQNVAVEDMEREVMEEMGESLTQII 931

RESULT 9
PCT-US01-01329-1409
; Sequence.1409, Application PC/TUS0101329
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al..
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA120PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01329
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 3506
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1409
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-01329-1409

Query Match                27.1%; Score 446; DB 1; Length 147;
Best Local Similarity     98.9%; Pred. No. 6.9e-34;
Matches    87; Conservative    0; Mismatches    1; Indels    0; Gaps    0;

QY   1 MVICQLKGGTQMLCIDNSRTELKALHLVPQYQQNNYLQSDVPKPMTALVGRLFPASTK 60
      |||||
Db   60 MVICQLKGGTQMLCIDNSRTELKALHLVPQYQQNNYLQSDVPKPMTALVGRLFPASTK 119

QY   61 LNLTQOLEGALPSVWNGSAFPGSGTLP 88
      |||||
Db   120 LNLTQOLEGALPSVWNGSAFPGSGTLP 147

RESULT 10
PCT-US01-01339-3988
; Sequence 3988, Application PC/TUS0101339
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc., et al..
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006PCT
; CURRENT APPLICATION NUMBER: PCT/US01/01339
; CURRENT FILING DATE: 2001-03-17
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3988
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (146)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
PCT-US01-01339-3988

Query Match                27.1%; Score 446; DB 1; Length 147;
Best Local Similarity     98.9%; Pred. No. 6.9e-34;
Matches    87; Conservative    0; Mismatches    1; Indels    0; Gaps    0;

QY   1 MVICQLKGGTQMLCIDNSRTELKALHLVPQYQQNNYLQSDVPKPMTALVGRLFPASTK 60
      |||||
Db   60 MVICQLKGGTQMLCIDNSRTELKALHLVPQYQQNNYLQSDVPKPMTALVGRLFPASTK 119

QY   61 LNLTQOLEGALPSVWNGSAFPGSGTLP 88
      |||||
Db   120 LNLTQOLEGALPSVWNGSAFPGSGTLP 147

```

```
Query Match      26.58; Score 437; DB 14; Length 438;
Best Local Similarity 41.3%; Pred. No. 2.2e-32;
Matches 92; Conservative 35; Mismatches 62; Indels 34; Gaps 8;
```

Qy	98	YCDCFASGDFENCCNCCNNLHHDIERFKAIRACLGRNPEAFQPKTG--KGQLGNVKP	155
		: : : : : :	
Db	191	YCDCFANGFECDNCDCNDCHNIIEVDQSRAIRQSLERNPNAPKPGIARGGITDIER	250
		: : : : : :	
Qy	156	QHNGKCNRSGGLKNYCCEYAQTMCSSICKTCGCKN-----YEES-----PERKTL	203
		: : : : : :	
Db	251	LHOKGHCHKSGCLKNYCCEYAAVPTDRCKCGCONTEYRMYTRKNSGGAVSNTNAL	310
		: : : : : :	

RESULT 15

Query Match 26.5%; Score 437; DB 16; Length 438;

Best Local Similarity	41.3%;	Pred. No.	2.2e-32;
Matches	92:	Conservative	35:
Mismatches	62:	Indels	34:
Gaps	8:		

Search completed: July 11, 2002, 08:20:36
Job time: 420 sec

; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6932
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-6932

Query Match 16.9%; Score 278; DB 6; Length 53;
Best Local Similarity 100.0%; Pred. No. 6.3e-17;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 VEATCACLAAQGEAEKEHCKSLAEQMIIEFGRCLSQILHTEFKSKGLKME 299
|||||
Db 1 VEATCACLAAQGEAEKEHCKSLAEQMIIEFGRCLSQILHTEFKSKGLKME 53

RESULT 6

US-10-155-881-20031
; Sequence 20031, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 20031
; LENGTH: 896
; TYPE: PRT
; ORGANISM: Glycine max
US-10-155-881-20031

Query Match 15.1%; Score 248.5; DB 6; Length 896;
Best Local Similarity 40.7%; Pred. No. 5.3e-13;
Matches 50; Conservative 13; Mismatches 31; Indels 29; Gaps 5;

QY 98 YDCDFASGDFCANN-CNCNCCNLLHDIERFRAIKACL-GRNPEAFQPKIGKQGLGNVKP 155
|||||
Db 487 YDCDFAAGTYCTDPCACQCGCLNRPEY-VETVYETKQOIESRNPFIAPAKI-----VQP 538
QY 156 -----QHNGKNCRRSGCLKNYCEYEAQIMCSTCKIGCKNYEE 196
|||
Db 539 TTDISHMDDENLTTPSSARHRRGCNCKRSMCLKKYCECYQANVGCGSCRCGCKNVHG 598

QY 197 SPE 199
|
Db 599 KXE 601

RESULT 7

US-10-155-881-9183
; Sequence 9183, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22

; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9183
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9183

Query Match 12.2%; Score 201.5; DB 6; Length 359;
Best Local Similarity 23.1%; Pred. No. 1.8e-09;
Matches 63; Conservative 35; Mismatches 102; Indels 73; Gaps 9;

QY 98 YDCDFASGDFCANN-CNCNCCN-----NLHHD-----IERFKAIKACLGRNPEAFQPKIG 146
|||||
Db 80 YPCFSGGGYCSKCCQCPFKKEAFETVHTTRKVLRSQKMSKINRRPEANTEPME 139
QY 147 KQQLGNVKPQHNGKNCRRSGCLKNYCEYEAQIMCSTCKIGCKNYEESPRTKTLMSM 206
|||||
Db 140 DAHSSSTTPPKRGCKCKSSCLKKYCDYQDGTGCSLFCRCDDCQNPFGKNEGIMADDS 199
QY 207 PNYMTGG-----LEGS-----HYLPP-----TK 225
|||
Db 200 KRYLYTGADLDHSEGEHDFVVERSPRLQSPISKESFHQTPPHLRASSRDAHVFPFAISO 259
QY 226 FSGLPFRSH--DRRPPSSCIS-----WEVVE-----ATCACLAAQGEAEKE 264
|||||
Db 260 WQALPRSHWCHSKNRKSSDRAMDDSNANYKSNHDYQVPHKEDSYSSKCVQILNGMAELS 319
QY 265 HCKSLAEQMIIEFGRCLSQILHTEFKSKGLK 297
|||||
Db 320 QVEKSVAPDVFLQPGNREIFVSLSGDVRAMWLK 352

RESULT 8

US-10-155-881-9877
; Sequence 9877, Application US/10155881
; GENERAL INFORMATION:
; APPLICANT: Dotson, Stanton B.
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; APPLICANT: Lutfiyya, Linda L.
; APPLICANT: McIninch, James
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(15300)J
; CURRENT APPLICATION NUMBER: US/10/155,881
; CURRENT FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 37595
; SEQ ID NO 9877
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Zea mays
US-10-155-881-9877

Query Match 11.1%; Score 183; DB 6; Length 143;
Best Local Similarity 40.9%; Pred. No. 2.4e-08;
Matches 36; Conservative 16; Mismatches 24; Indels 12; Gaps 3;

QY 98 YDCDFASGDFCANN-CNCNCCNLLHDIERFKAIKACLGRNPEAFQPKIGK-QGLGNV--- 153
|||||
Db 55 YCECFQELQYCDGCGNCSNCGNIVGNENARNEAIEAIRQORNPSPAFQPKIGDNTLNVRKD 114
QY 154 -----XPQHNGKNCRRSGCLKNYCE 174
|||
Db 115 KFWSSPSRSRSPKPHKCHCKKS-CLKKYCD 141

RESULT 9

US-10-138-145-1127
; Sequence 1127, Application US/10138145
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen

```
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: Gibson, John Bryan
; APPLICANT: Norris, Michael Geoffrey
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions isolated from forage
; FILE REFERENCE: (1)....(120)
; CURRENT APPLICATION NUMBER: US/10/138,145
; CURRENT FILING DATE: 2002-05-01
; NUMBER OF SEQ ID NOS: 152
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1127
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Lolium perenne
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)....(120)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-138-145-1127

Query Match      10.1%; Score 167; DB 6; Length 120;
Best Local Similarity 37.4%; Pred. No. 4.4e-07;
Matches 40; Conservative 15; Mismatches 36; Indels 16; Gaps 8;

QY 98 YDCDFASGDFCNN-CNCCNCCNLLHDIETKAIKACLGRLNPEAFOPKI-----GKGQLGN 152
Db 16 YCEC-AAGYVCSEPCGCGCLNKPHEIV-LSTRQIEFRNPLA-APKVIRLSDAAOETQ 72

QY 153 VKP-----QHNKGCNRRSGCLNVCYCEYBAQIMCSSICIGCKN 193
Db 73 EDPNTPASARHKGNCCKKSLK-YCEYOGGVGLTNCRK-ECKN 117

RESULT 10
PCT-US02-01339-6
; Sequence 6, Application PC/TUS0201339
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: WALIA, Narinder K.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: YUE, Henry
; APPLICANT: THANGAVELU, Kavitha
; APPLICANT: XU, Yuming
; APPLICANT: ARVIZU, Chandra
; APPLICANT: WARREN, Bridget A.
; APPLICANT: YAO, Monique G.
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: KALLICK, Deborah A.
; APPLICANT: GANDHI, Aneena R.
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: KHAN, Farrah A.
; APPLICANT: LU, Yan
; APPLICANT: SWARNAKAR, Anita
; APPLICANT: RAMKUMAR, Javalaxmi
; APPLICANT: NGUYEN, Dannel B.
; APPLICANT: GRAUL, Richard
; APPLICANT: LU, Dyung Aina M.
; TITLE OF INVENTION: RECEPTORS AND MEMBRANE-ASSOCIATED PROTEINS
; FILE REFERENCE: PI-0346 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/01339
; CURRENT FILING DATE: 2002-01-16
; PRIOR APPLICATION NUMBER: 60/262,838; 60/271,196; 60/274,549; 60/334,179
; PRIOR FILING DATE: 2001-01-19; 2001-02-02; 2001-02-23; 2001-03-09; 2001-11-28
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PERL Program
```

```
; SEQ ID NO 6
; LENGTH: 1774
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7722591CD1
PCT-US02-01339-6

Query Match      6.8%; Score 112; DB 1; Length 1774;
Best Local Similarity 20.2%; Pred. No. 0.46;
Matches 51; Conservative 33; Mismatches 73; Indels 96; Gaps 15;

QY 86 TLPGPPIITLAGYCDCAFSGDFCNN---CN-CNNCC---NNLH-----H 122
Db 800 TLPDP-----CRLLSSPEACNQGACTWCHGACLSGDAQHLRGLCGGSPMPRSPE 852

QY 123 DIERFKAIAKACLGRLNPEAFQPKIGKQLGNVYKPOHNGKNCRRSGCL-----KNYCE 174
Db 853 ECRRLRTCSECLARHPRTLQ--GDGEAST--PRCKWCTNCPGACIGRNGSCTSEND 908

QY 175 CYEAQIMCSSIC-----KCIGCKNYEESPERKTLMSM-PNY----- 209
Db 909 INQREVEFWAGNCSEACGAADCEQCTREGKCMWTRQFKRTGETRRLSVQPTDWT 968

QY 210 ---MOTGGLEGSHYLP-PTKFSGLPRFSDRRPSSCISWEVVEATCACLAAOGEAEKE 264
Db 969 SLLNVSPMPVSESPPLPCPTCHLLP-----NCTS-----CLDSKADGQWQ 1010

QY 265 HC-----SKCLA 271
Db 1011 HCVWSSSLQOCLS 1023

RESULT 11
PCT-US02-09671-718
; Sequence 718, Application PC/TUS0209671
; GENERAL INFORMATION:
; APPLICANT: ZYCOS Inc.
; TITLE OF INVENTION: TRANSLATIONAL PROFILING
; FILE REFERENCE: 08191-026W01
; CURRENT APPLICATION NUMBER: PCT/US02/09671
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: 60/279,495
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: 60/292,544
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/310,801
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/326,370
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/336,780
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/358,985
; PRIOR FILING DATE: 2002-02-20
; NUMBER OF SEQ ID NOS: 2041
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 718
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-09671-718

Query Match      6.6%; Score 109.5; DB 1; Length 699;
Best Local Similarity 22.3%; Pred. No. 0.26;
Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVICQLKGGTQMLCIDNSRTRRELKALHLVPOYQDQNNYVLOSVDPKPMTALVGRFLPASTK 60
Db 436 IVTVQVLPQCECRCDQSDRSL--CH-----GKGFLECGICRCDTGYIGKNCEQTQ 486
```


Job time: 40 sec

Best Local Similarity 22.3%; Pred. No. 0.29; Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVICQLKGGTQMLCIDNSRTRFKALHLVPOYQDNNYLOSDVPKPMYALVGRFLPASTK 60
 Db 436 IVTVQVLPQCECRCDQSRDRL--CH-----GKGFLEGGICRCDTGYIGNKCECQTQ 486
 QY 61 LNLITQOLEGA-----LPSVVNGSAFPGSTLPKPKITLAGYCDCEFA----- 103
 Db 487 -GRSSQLEGGSCRDNNNSIICSGGLGDCVCGOCLHTSDVPG--KLIYQYCECDTINCER 543
 QY 104 -SGDFCENN-----CNCNCCNHLHDIERF-----KAIKACLGRLNPEAFQPKIGKGOLGN 152
 Db 544 YNGQVCGGPGRLGFCGKC--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR--- 595
 QY 153 VKPOHNGCNCRRSGCLKNKYCEYEAQIM-----CSSIC-----KCIGCKNYEESP 198
 Db 596 -----CRC-----NVCECHSGYQLPLCQPCGCPSPCGKYISCAECLKEKGP 638

RESULT 15

PCT-US02-09671-712
 ; Sequence 712, Application PC/TUS0209671
 ; GENERAL INFORMATION:
 ; APPLICANT: Zycos Inc.
 ; TITLE OF INVENTION: TRANSLATIONAL PROFILING
 ; FILE REFERENCE: 08191-026W01
 ; CURRENT APPLICATION NUMBER: PCT/US02/09671
 ; CURRENT FILING DATE: 2002-03-28
 ; PRIOR APPLICATION NUMBER: 60/219,495
 ; PRIOR FILING DATE: 2001-03-28
 ; PRIOR APPLICATION NUMBER: 60/292,544
 ; PRIOR FILING DATE: 2001-05-21
 ; PRIOR APPLICATION NUMBER: 60/310,801
 ; PRIOR FILING DATE: 2001-08-08
 ; PRIOR APPLICATION NUMBER: 60/326,370
 ; PRIOR FILING DATE: 2001-10-01
 ; PRIOR APPLICATION NUMBER: 60/336,780
 ; PRIOR FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 60/358,985
 ; PRIOR FILING DATE: 2002-02-20
 ; NUMBER OF SEQ ID NOS: 2041
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 712
 ; LENGTH: 769
 ; TYPE: prt
 ; ORGANISM: Homo sapiens
 PCT-US02-09671-712

Query Match 6.6%; Score 109.5; DB 1; Length 769;
 Best Local Similarity 22.3%; Pred. No. 0.29; Matches 53; Conservative 26; Mismatches 84; Indels 75; Gaps 15;

QY 1 MVICQLKGGTQMLCIDNSRTRFKALHLVPOYQDNNYLOSDVPKPMYALVGRFLPASTK 60
 Db 436 IVTVQVLPQCECRCDQSRDRL--CH-----GKGFLEGGICRCDTGYIGNKCECQTQ 486
 QY 61 LNLITQOLEGA-----LPSVVNGSAFPGSTLPKPKITLAGYCDCEFA----- 103
 Db 487 -GRSSQLEGGSCRDNNNSIICSGGLGDCVCGOCLHTSDVPG--KLIYQYCECDTINCER 543
 QY 104 -SGDFCENN-----CNCNCCNHLHDIERF-----KAIKACLGRLNPEAFQPKIGKGOLGN 152
 Db 544 YNGQVCGGPGRLGFCGKC--RCHPGFEGSACQERTTEGCL--NPRRVECS-GRGR--- 595
 QY 153 VKPOHNGCNCRRSGCLKNKYCEYEAQIM-----CSSIC-----KCIGCKNYEESP 198
 Db 596 -----CRC-----NVCECHSGYQLPLCQPCGCPSPCGKYISCAECLKEKGP 638

Fri Jul 12 08:35:31 2002

us-09-743-237-5.rapn

Page 7

